

586

```
Db 271 GWCAINSKILVPPRTTHAKFTDWKYLKRLVGSRTLPVDFHKKVSMKIPFQGMRELE 330
Qy 121 VVDKSOVSRTRMAVVDVTVIGRLRLLYEDGSDDDFWCHMWSPLIHPVGSRRVGHGIGM 180
Db 331 VVDKSOVSRTRMAVVDVTVIGRLRLLYEDGSDDDFWCHMWSPLIHPVGSRRVGHGIGM 390
Qy 181 SERSDMAHHTPRKIYCDVPIYFKKRVAVYTGSGWFEEGKMLEADPLNLGNICVATV 240
Db 391 SERSDMAHHTPRKIYCDVPIYFKKRVAVYTGSGWFEEGKMLEADPLNLGNICVATV 450
Qy 241 CKVLLDGYLMICVDGGSTGDLDFCYHASSHAIFPATFCQKNDIELTPPKGYEAQTFNW 300
Db 451 CKVLLDGYLMICVDGGSTGDLDFCYHASSHAIFPATFCQKNDIELTPPKGYEAQTFNW 510
Qy 301 ENYLEKTSKAAPRLFNMDCPNHPGKVMKLEAVDLMEPRLLICVATVKKVVRHLLSIHF 360
Db 511 ENYLEKTSKAAPRLFNMDCPNHPGKVMKLEAVDLMEPRLLICVATVKKVVRHLLSIHF 570
Qy 361 DGWSEYDQWVDCESPDIPYVGCWELTGYLOPPVAAEPATPLKAKEATKKKKQFGKKR 420
Db 571 DGWSEYDQWVDCESPDIPYVGCWELTGYLOPPVAAEPATPLKAKEATKKKKQFGKKR 630
Qy 421 KRIPPTKTRPLRQSGKKPLLEDDPQGARKISSEVPVGEIIIAVRVKEEHLDVASPDKASSP 480
Db 631 KRIPPTKTRPLRQSGKKPLLEDDPQGARKISSEVPVGEIIIAVRVKEEHLDVASPDKASSP 690
Qy 481 ELPVSVENIKQETDD 495
Db 591 ELPVSVENIKQETDD 705
```

RESULT 2

```
IML2 MOUSE
ID _IML2_MOUSE STANDARD; PRT: 703 AA.
EC PS9178;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lethal(3)malignant brain tumor-like 2 protein (L(3)mbt-like 2 protein)
DE (H-1(3)mbt-like protein).
GN L3MBTL2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Liver;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares J.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: Putative Polycomb group (PcG) protein. PcG proteins
CC maintain the transcriptionally repressive state of genes, probably
CC via a modification of chromatin, rendering it heritably changed in
```

```
CC its expressibility. Its association with a chromatin remodeling
CC complex suggests that it may contribute to prevent expression of
CC genes that trigger the cell into mitosis (By similarity).
CC -1- SUBUNIT: Part of the E2F6.com-1 complex in G0 phase composed of
CC E2F6, MGA, MAX, TFDPl, CBX3, BATF, E2F6, RING1, RNF2, MBLR,
CC BATF and YAF2 (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: Contains 1 C4-type zinc finger.
CC -1- SIMILARITY: Contains 4 MBT domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; BC030864; AAH30864.1;
CC MGD; MGI:2443584; 4732493N06Rik.
CC SMART; SM00561; MBT; 4.
CC Transcription regulation; Chromatin regulator; Nuclear protein;
CC Zinc-finger; Metal-binding; Repeat.
CC ZN_FING 90 114 C4-TYPE.
CC FT DOMAIN 214 291 MBT 1.
CC FT DOMAIN 327 389 MBT 2.
CC FT DOMAIN 432 509 MBT 3.
CC FT DOMAIN 540 611 MBT 4.
CC FT DOMAIN 17 21 POLY-GLU.
CC FT DOMAIN 620 624 POLY-LYS.
CC SQ SEQUENCE 703 AA; 78954 MW; 259B5DF5BFDAAE9 CRC64;
```

```
Query Match 90.8%; Score 2436; DB 1; Length 703;
Best Local Similarity 89.7%; Pred. No. 1.6e-195;
Matches 444; Conservative 23; Mismatches 26; Indels 2; Gaps 1;
Qy 1 MKGMKEVLNSDAVLPSRVYVIAVOTAGYRVLRLVREGFENDASHDFWCLNLTVDVHPI 60
Db 211 MKGMKEVLNSDAVLPSRVYVIAVOTAGYRVLRLVREGFENDASHDFWCLNLTVDVHPI 270
Qy 61 GWCAINSKILVPPRTTHAKFTDWKYLKRLVGSRTLPVDFHKKVSMKIPFQGMRELE 120
Db 271 GWCAINSKILVPPRTTHAKFTDWKYLKRLVGSRTLPVDFHKKVSMKIPFQGMRELE 330
Qy 121 VVDKSOVSRTRMAVVDVTVIGRLRLLYEDGSDDDFWCHMWSPLIHPVGSRRVGHGIGM 180
Db 331 VVDKSOVSRTRMAVVDVTVIGRLRLLYEDGSDDDFWCHMWSPLIHPVGSRRVGHGIGM 390
Qy 181 SERSDMAHHTPRKIYCDVPIYFKKRVAVYTGSGWFEEGKMLEADPLNLGNICVATV 240
Db 391 SERSDMAHHTPRKIYCDVPIYFKKRVAVYTGSGWFEEGKMLEADPLNLGNICVATV 450
Qy 241 CKVLLDGYLMICVDGGSTGDLDFCYHASSHAIFPATFCQKNDIELTPPKGYEAQTFNW 300
Db 451 CKVLLDGYLMICVDGGSTGDLDFCYHASSHAIFPATFCQKNDIELTPPKGYEAQTFNW 510
Qy 301 ENYLEKTSKAAPRLFNMDCPNHPGKVMKLEAVDLMEPRLLICVATVKKVVRHLLSIHF 360
Db 511 ENYLEKTSKAAPRLFNMDCPNHPGKVMKLEAVDLMEPRLLICVATVKKVVRHLLSIHF 570
Qy 361 DGWSEYDQWVDCESPDIPYVGCWELTGYLOPPVAAEPATPLKAKEATKKKKQFGKKR 420
Db 571 DGWSEYDQWVDCESPDIPYVGCWELTGYLOPPVAAEPATPLKAKEATKKKKQFGKKR 630
Qy 421 KRIPPTKTRPLRQSGKKPLLEDDPQGARKISSEVPVGEIIIAVRVKEEHLDVASPDKASSP 480
Db 631 KRIPPTKTRPLRQSGKKPLLEDDPQGARKISSEVPVGEIIIAVRVKEEHLDVASPDKASSP 690
Qy 481 ELPVSVENIKQETDD 495
Db 591 ELPVSVENIKQETDD 705
```

240
AAGA
|||
AAGA 874

RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tanlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kerhaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA Martin I.D., Mashreghi-Mohammadi M., Matthews L.H., McCann O.T.,
RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.E., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.B., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Shibuya K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyma S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Slean D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Rao Z.,
RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Benis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA Emanuel B.S., Shaith T., Kurahashi H., Saitta S., Budarf M.L.,
RA McDermid H.E., Johnson A., Wong A.C.C., Morrow S.E., Edelmann L.,
RA Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M.K., Kedra D.,
RA Serousi E., Franssen I., Tapia I., Bruder C.E., O'Brien K.P.,
RA Wilkinson P., Bodetrich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tilahun Y., Wright H.,
RL "The DNA sequence of human chromosome 22";
RL Nature 402:489-495 (1999).
RN (5).
RN SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Placenta;
RX MEDLINE=23898257; PubMed=12477932;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RX Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RX Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RX Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RX Diatchenko L., Narusina K., Farmer A.A., Rubin G.M., Hong L.,
RX Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RX Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RX Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RX Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RX Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RX Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RX Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RX Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RX Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RX Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RX Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RX Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN (6).
RP IDENTIFICATION OF COMPLEX WITH B2F6; TFDPI; MAX; MGA; EUHMTASE1; CBX3;
RP RING1; RNF2; MBLR; BAF8 AND YAF2.
RX MEDLINE=2199559; PubMed=12004135;
RX Ogawa H., Ishiguro K.-I., Gaubatz S., Livingston D.M., Nakatani Y.,
RT "A complex with chromatin modifiers that occupies B2F- and
RT Myc-responsive genes in GO cells";
RL Science 296:1132-1136 (2002).
CC -1- FUNCTION: Putative Polycomb group (PcG) protein. PcG proteins
CC maintain the transcriptionally repressive state of genes, probably
CC via a modification of chromatin, rendering it heritably changed in
CC its expressibility. Its association with a chromatin remodeling

complex suggests that it may contribute to prevent expression of
genes that trigger the cell into mitosis.
-1- SUBUNIT: Part of the E2F6-com-1 complex in G0 phase composed of
B2F6, MGA, MAX, TFDPI, CBX3, BAF8, EUHMTASE1, RING1, RNF2, MBLR,
BAF8 and YAF2.
-1- SUBCELLULAR LOCATION: Nuclear (Probable).
-1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=3;
Name=1; Synonyms=A;
IsoId=Q969R5-1; Sequence=Displayed;
Name=2; Synonyms=B;
IsoId=Q969R5-2; Sequence=VSP_003904; VSP_003905;
Name=3;
IsoId=Q969R5-3; Sequence=VSP_003906; VSP_003907;
-1- SIMILARITY: Contains 1 C4-type zinc finger.
-1- SIMILARITY: Contains 4 MBT domains.
-1- CAUTION: Ref.3 sequences differ from that shown in that they seem
to include intronic sequence.
-1- CAUTION: Ref.4 sequence differs from that shown due to erroneous
gene model prediction.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

CC EMBL: AJ305226; CAC37794.1; -
CC EMBL: AJ305227; CAC37795.1; -
CC EMBL: AL136564; CAB6499.1; -
CC EMBL: AK074091; BAB84917.1; ALT SEQ.
CC EMBL: AK097052; BAC04936.1; ALT SEQ.
CC EMBL: AL035658; -; NOT ANNOTATED CDS.
CC EMBL: AL035681; CAB63071.1; ALT SEQ.
CC EMBL: BC017191; AAH17191.1; -
CC Genew; HGNC:18594; L3MBTL2.
CC GK; Q969R5; -
CC InterPro; IPR004092; MBT.
CC Pfam; PF02820; mbt; 4.
CC SMART; SM00561; MBT; 4.
KW Transcription regulation; Chromatin regulator; Nuclear protein;
KW Zinc-finger; Metal-binding; Repeat; Alternative splicing;
KW Polymorphism. 90 114 C4-TYPE.
FT ZN_FING 214 291 MBT 1.
FT DOMAIN 327 399 MBT 2.
FT DOMAIN 432 509 MBT 3.
FT DOMAIN 540 611 MBT 4.
FT DOMAIN 17 20 POLY-GLU.
FT DOMAIN 620 624 POLY-LYS.
FT VARSPPLIC 608 614 EPATPLK -> GKLPRL (in isoform 2).
FT VARSPPLIC 615 705 Missing (in isoform 2).
FT VARSPPLIC 608 617 /FTid=VSP_003905.
FT VARSPPLIC 618 705 EPATPLAKE -> GVGRGPKRL (in isoform 3).
FT VARSPPLIC 618 705 /FTid=VSP_003906.
FT VARIANT 300 300 Missing (in isoform 3).
FT VARIANT 300 300 R -> W (IN dSNP:2277846).
FT SEQUENCE 705 AA; 79110 MW; 8FC86A440982FFA7 CRC64;
Query Match 100.0%; Score 2683; DB 1; Length 705;
Best Local Similarity 100.0%; Pred. No. 3.8e-216;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKGMKEVLNDAVLPSRVYVIAVSIQTAGRVLLRYEGFENDASHDFWNCNLCTVDVHPI 60
DB 211 MKGMKEVLNDAVLPSRVYVIAVSIQTAGRVLLRYEGFENDASHDFWNCNLCTVDVHPI 270
QY 61 GWCAINSKILVPPRTTHAKFTDWKGYLMKVLGSRTLPLVDFTHKMYESMKYPFQGMRL 120

FT:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2004, 09:05:51 ; Search time 18 Seconds
(without alignments)
1293.234 Million cell updates/sec

Title: US-10-031-915-36

Perfect score: 2683

Sequence: 1 MKGMKVEVLNSDAVLPSRVY.....KASSPELPVSVENIKQETDD 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2683	100.0	705	1 LML2_HUMAN	Q969r5 homo sapien
2	2436	90.8	703	1 LML2_MOUSE	P59178 mus musculus
3	510	19.0	772	1 LMBT_HUMAN	Q9Y468 homo sapien
4	121	4.5	867	1 SSPO_BOVIN	P98167 bos taurus
5	116.5	4.3	200	1 RL29_MYCGE	P47405 mycoplasma
6	109.5	4.1	1026	1 TAC2_HUMAN	Q95359 homo sapien
7	104	3.9	6669	1 NEB1_HUMAN	P20329 homo sapien
8	98.5	3.7	738	1 PAP_BOVIN	P25500 bos taurus
9	98.5	3.7	738	1 PAP_MOUSE	Q61183 mus musculus
10	96.5	3.6	1377	1 Z198_HUMAN	Q9ubw7 homo sapien
11	95.5	3.6	721	1 ENP1_TORCA	P14400 torpedo cal
12	92.5	3.4	479	1 CBP5_CANAL	O43101 candida alb
13	92.5	3.4	1593	1 AT12_HUMAN	P58397 homo sapien
14	91.5	3.4	245	1 H1_WAIZE	P23444 zea mays (m
15	91	3.4	401	1 EAF3_YEAST	Q12432 saccharomyc
16	91	3.4	486	1 H51_HUMAN	P14317 homo sapien
17	90.5	3.4	2150	1 SDC3_CAREL	P34706 caenorhabdi
18	90	3.4	520	1 GAG_SIVAI	P27972 simian immu
19	90	3.4	1020	1 NFH_HUMAN	P12036 homo sapien
20	89.5	3.3	510	1 CP50_CANMA	Q12587 candida mal
21	89	3.3	3924	1 ANK2_MOUSE	Q10484 homo sapien
22	88.5	3.3	931	1 NRP2_MOUSE	O35375 mus musculus
23	88	3.3	202	1 H1_LYCPN	P40267 lycopersico
24	88	3.3	2459	1 MAPB_RAT	P15205 rattus norv
25	87.5	3.3	425	1 SMA3_HUMAN	Q92940 homo sapien
26	87.5	3.3	925	1 NRP2_RAT	O35276 rattus norv
27	87	3.2	519	1 GAG_SIVAT	P05892 simian immu
28	86.5	3.2	307	1 VP53_BPAPS	Q9tp5 bacterioph
29	86.5	3.2	560	1 PTK1_YEAST	P36002 saccharomyc
30	86.5	3.2	651	1 CORO_YEAST	Q06440 saccharomyc
31	86.5	3.2	1714	1 SYEP_DROME	P28668 drosophila
32	86.5	3.2	2434	1 ABC2_MOUSE	P41234 mus musculus
33	86	3.2	400	1 PD12_MORAP	Q9Y8H5 mortierella

34	86	3.2	1156	1 JAK1_CYPCA	Q09178 cyprinus ca
35	86	3.2	1406	1 TOPI_CANGA	Q93794 candida gla
36	86	3.2	1440	1 SYEP_HUMAN	P07814 homo sapien
37	85.5	3.2	414	1 YAF4_ECOLI	P04335 escherichia
38	85.5	3.2	876	1 SYA_THEVO	Q97ah7 thermoplas
39	85	3.2	400	1 FD13_MORIS	P59668 mortierella
40	85	3.2	484	1 PAP3_XENLA	P51005 xenopus lae
41	85	3.2	600	1 NUCD_BUCAL	P27254 buchnera ap
42	85	3.2	601	1 GLMS_PYRAE	Q8ztz0 p glucosami
43	85	3.2	619	1 PFCK_HAEFO	P29190 haemochus
44	85	3.2	1075	1 NFK3_HUMAN	Q12968 homo sapien
45	85	3.2	2468	1 MAPB_HUMAN	P46821 homo sapien

ALIGNMENTS

RESULT 1					
LML2_HUMAN					
ID	LML2_HUMAN	STANDARD;	PRT;	705 AA.	
AC	Q969r5; Q9TEN1; Q96SC4; Q9BQ12; Q9UGS4;				
DT	28-FEB-2003 (Rel. 41, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DT	15-SEP-2003 (Rel. 42, Last annotation update)				
DE	Lethal(3)malignant brain tumor-like 2 protein (L(3)mbt-like 2 protein)				
DE	(H-1(3)mbt-like protein).				
GN	L3MBTL2				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).				
RP	MEDLINE=21538645; PubMed=11692070;				
RA	Wisnar J.;				
RT	"Molecular characterization of h-l(3)mbt-like: a new member of the				
RT	human mbt family.";				
RL	FEBS Lett. 507:119-121(2001).				
RN	[2]				
RP	SEQUENCE FROM N.A. (ISOFORM 3).				
RC	TISSUE=Amysdala;				
RC	MEDLINE=21154917; PubMed=11230166;				
RA	Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,				
RA	Ansorge W., Boecker M., Bloecker H., Bauersachs S., Blum H.,				
RA	Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,				
RA	Mewes H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,				
RA	Wambutt R., Korn B., Klein M., Poustka A.;				
RT	"Towards a catalog of human genes and proteins: sequencing and				
RT	analysis of 500 novel complete protein coding human cDNAs.";				
RL	Genome Res. 11:422-435(2001).				
RN	[3]				
RP	SEQUENCE FROM N.A. (ISOFORM 3).				
RC	TISSUE=Small intestine, and Spleen;				
RA	Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,				
RA	Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,				
RA	Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,				
RA	Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,				
RA	Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,				
RA	Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,				
RA	Kawakami B., Sugihara K., Sugano S., Nagahari K., Masuho Y., Nagai K.,				
RA	Isogai T., Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T.,				
RA	Ohara O.;				
RT	"NEDO human cDNA sequencing project.";				
RL	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RP	MEDLINE=20057165; PubMed=10591208;				
RA	Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,				
RA	Clamp M., Smak L.J., Ainscough R., Almeida J.P., Babbage A.K.,				
RA	Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,				
RA	Bird C.P., Blake S.E., Bridgman A.M., Buck D., Burgess J.,				
RA	Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,				
RA	Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,				

QY 661 GAGCAGTCTACACAGAGGCGGTTGGTTTGGAGAGGGATGAAGCTGGAGGCCATTGACC 720
 Db 1295 GAGCAGTCTACACAGAGGCGGTTGGTTTGGAGAGGGATGAAGCTGGAGGCCATTGACC 1354
 QY 721 CCCTGAATCTGGGCAACATCTGCGTGGCAATCTGTAAGGTTCTCTGGATGGATACC 780
 Db 1355 CCCTGAATCTGGGCAACATCTGCGTGGCAATCTGTAAGGTTCTCTGGATGGATACC 1414
 QY 781 TGATGATCTGTGTGGAGCGGGGCGCTCCACAGATGGCTTGGACTGTCTGTACATG 840
 Db 1415 TGATGATCTGTGTGGAGCGGGGCGCTCCACAGATGGCTTGGACTGTCTGTACATG 1474
 QY 841 CCTCTTCCAGCCCATCTTCCCGGCCACTTCTGTGAGAGATGACATTCAGCTCAC 900
 Db 1475 CCTCTTCCAGCCCATCTTCCCGGCCACTTCTGTGAGAGATGACATTCAGCTCAC 1534
 QY 901 CGCCAAAGGTTATGAGGCAAGACTTCAACTGGGAGAACTACTTGGAGAGACCAAGT 960
 Db 1535 CGCCAAAGGTTATGAGGCAAGACTTCAACTGGGAGAACTACTTGGAGAGACCAAGT 1594
 QY 961 CGAAAGCGCTCCATCGAGCTCTTTAACTGATGATGCGCAACCATGGCTTCAAGTGG 1020
 Db 1595 CGAAAGCGCTCCATCGAGCTCTTTAACTGATGATGCGCAACCATGGCTTCAAGTGG 1654
 QY 1021 GCATGAGCTGGAGCGGCTGAGCTGATGAGAGCGCCGCTCATCTGTGTGCGCAAGTGA 1080
 Db 1655 GCATGAGCTGGAGCGGCTGAGCTGATGAGAGCGCCGCTCATCTGTGTGCGCAAGTGA 1714
 QY 1081 AACGAGTGGTGCATCGGCTCTCAGCTGATGATGAGAGCGGCTGAGAGCGGCTGAGTGA 1140
 Db 1715 AACGAGTGGTGCATCGGCTCTCAGCTGATGATGAGAGCGGCTGAGAGCGGCTGAGTGA 1774
 QY 1141 AGTGGTGGATCGAGTCCCGAGATCTACCTGAGAGCGGCTGAGTGGTGGTGGTGGTGGT 1200
 Db 1775 AGTGGTGGATCGAGTCCCGAGATCTACCTGAGAGCGGCTGAGTGGTGGTGGTGGTGGT 1834
 QY 1201 ACCAGTCCAGCTCTCTGCGCGAGAGCGGCGCAACCGCTGAGAGCGGCTGAGAGCGGCTGAG 1260
 Db 1835 ACCAGTCCAGCTCTCTGCGCGAGAGCGGCGCAACCGCTGAGAGCGGCTGAGAGCGGCTGAG 1894
 QY 1261 CAAGAAGAAAGAAACAGTTTGGAGAGAAAGAAAGAAATCCCGCCCACTAAGACGC 1320
 Db 1895 CAAGAAGAAAGAAACAGTTTGGAGAGAAAGAAAGAAATCCCGCCCACTAAGACGC 1954
 QY 1321 GACCCCTCAGACAGGGTCCAGAGAGCGGCTGAGAGAGCGGCTGAGAGAGCGGCTGAGAGAG 1380
 Db 1955 GACCCCTCAGACAGGGTCCAGAGAGCGGCTGAGAGAGCGGCTGAGAGAGCGGCTGAGAGAG 2014
 QY 1381 AGATCTCGTGGAGCGCTGTTCTGCGGAGATCATGCTGCGTGTGAAGAGAGCATC 1440
 Db 2015 AGATCTCGTGGAGCGCTGTTCTGCGGAGATCATGCTGCGTGTGAAGAGAGCATC 2074
 QY 1441 TAGAGTGGCTGCGCGAGAGGTTCAAGTCCAGAGCTGCTGCTGCTGCGTCCAGAGCA 1500
 Db 2075 TAGAGTGGCTGCGCGAGAGGTTCAAGTCCAGAGCTGCTGCTGCTGCGTCCAGAGCA 2134
 QY 1501 TCAAGCAGGAAACAGACGACTGAGCTTCTGCGCTCCAGAGCTGCTGCTGCTGCGTCCAGAG 1560
 Db 2135 TCAAGCAGGAAACAGACGACTGAGCTTCTGCGCTCCAGAGCTGCTGCTGCTGCGTCCAGAG 2194
 QY 1561 AGCCAGCGGTTTCTTACACACACACATGCTTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 Db 2195 AGCCAGCGGTTTCTTACACACACACATGCTTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCT 2254
 QY 1621 TCCTCTCTGTGTAATTTCTGCGCGGCTGCTGAGAGCTGAGAGCTGAGAGCTGCTGCTGCTGCT 1680
 Db 2255 TCCTCTCTGTGTAATTTCTGCGCGGCTGCTGAGAGCTGAGAGCTGAGAGCTGCTGCTGCTGCT 2314
 QY 1681 GTCTCTCTGAGACCGGCTGCTGCTTCTGCGCTTCTGCGCTTCTGCGCTTCTGCGCTTCTGCG 1740
 Db 2315 GTCTCTCTGAGACCGGCTGCTGCTTCTGCGCTTCTGCGCTTCTGCGCTTCTGCGCTTCTGCG 2374
 QY 1741 CGCCTGAGGCGCCAGAACTGCTGCTGTGAACACCTTTTCCAGCCAGAGTTCCCAAGCTG 1800

Db 2375 CGCCTGAGGCCCCAGAACTGCTGCTGAAACCACTTTTCAGCCAGAGTTCCCAAGCTG 2434
 QY 1801 GAAAGCTAGCTGCTGCTGCTTCTTAAAGATGGCTCCCGCCGACCCCGCCAGCGGCTCAG 1860
 Db 2435 GAAAGCTAGCTGCTGCTGCTTCTTAAAGATGGCTCCCGCCGACCCCGCCAGCGGCTCAG 2494
 QY 1861 TTGCCAGGATGGGCGCCACACTGCTCACCTGTGGAATACAAAGACAGTGAATCTGTCTG 1920
 Db 2495 TTGCCAGGATGGGCGCCACACTGCTCACCTGTGGAATACAAAGACAGTGAATCTGTCTG 2554
 QY 1921 CTTGAACAGCTCATGTAATTAAGTTCTAGAGAGCTCTCTGAGCAGGATAAGTCCCT 1980
 Db 2555 CTTGAACAGCTCATGTAATTAAGTTCTAGAGAGCTCTCTGAGCAGGATAAGTCCCT 2614
 QY 1981 GACAGTGTGTGTGTGGGGGAGCTCTGCTCAAAAATTCACAGCAGATGCT 2040
 Db 2615 GACAGTGTGTGTGTGGGGGAGCTCTGCTCAAAAATTCACAGCAGATGCT 2674
 QY 2041 CTTGAGCTCATGTTGTGCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
 Db 2675 CTTGAGCTCATGTTGTGCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2734
 QY 2101 CTTGCTCGGAGCTTAAGAACAGCTGACAGAGATGTGATTTGGCGACCTGTGTGTGG 2160
 Db 2735 CTTGCTCGGAGCTTAAGAACAGCTGACAGAGATGTGATTTGGCGACCTGTGTGTGG 2794
 QY 2161 CTTGAGCTGCTTCTGT 2220
 Db 2795 CTTGAGCTGCTTCTGT 2854
 QY 2221 GGGGAGGACATTTGGAG 2280
 Db 2855 GGGGAGGACATTTGGAG 2914
 QY 2281 GAAGCCCGCTAAAAATTAATTCATCCAGATTCCTTTGTAGTAAAGGTCAGTTCTGA 2340
 Db 2915 GAAGCCCGCTAAAAATTAATTCATCCAGATTCCTTTGTAGTAAAGGTCAGTTCTGA 2974
 QY 2341 CTTGAGCTCTAGAGAGCTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2400
 Db 2975 CTTGAGCTCTAGAGAGCTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3034
 QY 2401 AAGAAACCATGCTCGAGGCGCGTGAACACAGAACCTCAAGACAGAGATGACAGCT 2460
 Db 3035 AAGAAACCATGCTCGAGGCGCGTGAACACAGAACCTCAAGACAGAGATGACAGCT 3094
 QY 2461 GGAGGACATCTAGCTGCCATTGCAAGCTCACTGGGCTCCCGACACTGTGTGTGTGAGA 2520
 Db 3095 GGAGGACATCTAGCTGCCATTGCAAGCTCACTGGGCTCCCGACACTGTGTGTGTGAGA 3154
 QY 2521 AATTAAACCCCTGCTTGTGCTGAGAAAAA 2555
 Db 3155 AATTAAACCCCTGCTTGTGCTGAGAAAAA 3189

RESULT 3

BC017191
 LOCUS 3222 bp mRNA linear PRI 09-NOV-2001
 DEFINITION Homo sapiens, hypothetical protein DKFZp761i1141, clone MGC:2476
 IMAGE:3138444, mRNA, complete cds.
 ACCESSION BC017191
 VERSION BC017191.1 GI:16877934
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 3222)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (09-NOV-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2004, 09:05:51 ; Search time 18 Seconds
(without alignments)
1293.234 Million cell updates/sec

Title: US-10-031-915-36

Perfect score: 2683

Sequence: 1 MKGMKVEVNSDAVLPSRVY.....KASSPELFSVENIKQETDD 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2683	100.0	705	1 LML2_HUMAN	Q969r5 homo sapien
2	2436	90.8	703	1 LML2_MOUSE	P59178 mus musculus
3	510	19.0	772	1 LMBT_HUMAN	Q9Y468 homo sapien
4	121	4.5	867	1 SSPB_BOVIN	P98167 bos taurus
5	116.5	4.3	200	1 RL29_MYCCE	P47405 mycoplasma
6	109.5	4.1	1026	1 TAC2_HUMAN	Q95359 homo sapien
7	104	3.9	6869	1 NBSU_HUMAN	P20529 homo sapien
8	98.5	3.7	738	1 PAP_BOVIN	P25500 bos taurus
9	98.5	3.7	738	1 PAP_MOUSE	Q61183 mus musculus
10	96.5	3.6	1377	1 Z198_HUMAN	Q9ubw7 homo sapien
11	95.5	3.6	721	1 ENP1_TORCA	P14400 torpedo cal
12	92.5	3.4	479	1 CBP5_CANAL	O43101 candida alb
13	92.5	3.4	1593	1 ATL2_HUMAN	P58397 homo sapien
14	91.5	3.4	245	1 H1_MALIZE	P23444 zea mays (m
15	91	3.4	401	1 EAF3_YEAST	Q12432 saccharomyc
16	91	3.4	486	1 HSI_HUMAN	P14317 homo sapien
17	90.5	3.4	2150	1 SDG3_CAEEL	P34706 caenorhabdi
18	90	3.4	520	1 GAG_SIVAI	P27972 simian immu
19	90	3.4	1020	1 NEH_HUMAN	P12036 homo sapien
20	89.5	3.3	510	1 CP5Q_CANMA	Q12587 candida mal
21	89	3.3	3924	1 ANK2_HUMAN	Q01484 homo sapien
22	88.5	3.3	931	1 NRP2_MOUSE	O35375 mus musculus
23	88	3.3	202	1 H1_LYCPFN	P40267 lycopersico
24	88	3.3	2459	1 MAPB_RAT	P15205 rattus norv
25	87.5	3.3	425	1 SMA3_HUMAN	Q92940 homo sapien
26	87.5	3.3	925	1 NRP2_RAT	O35276 rattus norv
27	87	3.2	519	1 GAG_SIVAT	P58992 simian immu
28	86.5	3.2	307	1 VP53_EPAFS	Q9tip5 bacterioph
29	86.5	3.2	560	1 PTK1_YEAST	P36002 saccharomyc
30	86.5	3.2	651	1 CORO_YEAST	Q05440 saccharomyc
31	86.5	3.2	1714	1 SYEP_DROME	P28668 drosophila
32	86.5	3.2	2434	1 ABC2_MOUSE	P41234 mus musculus
33	86	3.2	400	1 FD12_MORAP	Q9y8h5 mortierella

RESULT 1
LML2_HUMAN
ID LML2_HUMAN STANDARD; PRT; 705 AA.
AC Q969R5; Q8TEN1; Q96SC4; Q9BQI2; Q9UGS4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lethal(3)malignant brain tumor-like 2 protein (L(3)mbt-like 2 protein)
DE (H-1(3)mbt-like protein).
GN L3MBTL2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=21538645; PubMed=11682070;
RA Wismar J.;
RT "Molecular characterization of h-1(3)mbt-like: a new member of the
RL human mbt family.";
RL FEBS Lett. 507:119-121(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=amygdala;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansoerge W., Boecker M., Bloeker H., Bauersachs S., Blum H.,
RA Lauber J., Dueterhoeft A., Beyer A., Koshner K., Strack N.,
RA Mewes H.-W., Osterweider B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;
RT "Towards a catalog of human genes and proteins: sequencing and
RL analysis of 500 novel complete protein coding human cDNAs.";
RN Genome Res. 11:422-435(2001).
[3]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=small intestine, and spleen;
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA Isogai T., Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T.,
RA Ohara O.;
RT "NEDO human cDNA sequencing project.";
RN Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,
RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cooley V.E., Cole C.G., Collier R.E., Connor R.,

*

Q09178 cyprinus ca
Q93794 candida gla
P07814 homo sapien
P04335 escherichia
Q97ah7 thermoplasma
P59668 mortierella
P51005 xenopus lae
P57254 buchnera ap
Q8zt20 p glucosami
P29190 haemochus
Q12868 homo sapien
P46821 homo sapien

ALIGNMENTS

RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dhani P.D., Dockree C., Dodsorth S.J., Durbin R.M., Ellington A.G.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Gratham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kerhaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spragon L., Steward C.A., Suleston J.E., Swann R.M.,
RA Vaadin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kuch J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Leo H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Mala J.E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Kohlfing T.,
RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
RA Korf I., Bedell J.A., Hallier L., Mardis E., Waterston R., Wilson R.,
RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
RA McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
RA Kim U.J., Shizuya H., Simon M.I., Dumanaki J.P., Peyrard M., Kedra D.,
RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P., Lane L.,
RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tilahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUEPlacenta;
RX MEDLINE=2389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [6]

RP IDENTIFICATION OF COMPLEX WITH E2F6; TFDPI; MAX; MGA; EUMTASE1; CBX3;
RP RING1; RNF2; MBLR; BAT8 AND YAP2.
RX MEDLINE=2199559; PubMed=12004135;
RA Ogawa H., Ishiguro K.-I., Gaubatz S., Livingston D.M., Nakatani Y.;
RT "A complex with chromatin modifiers that occupies E2F- and
Myc-responsive genes in G0 cells.";
RL Science 296:1132-1136(2002).
CC -!- FUNCTION: Putative Polycomb group (PcG) protein. PcG proteins
maintain the transcriptionally repressive state of genes, probably
via a modification of chromatin, rendering it heritably changed in
its expressibility. Its association with a chromatin remodeling

complex suggests that it may contribute to prevent expression of
genes that trigger the cell into mitosis.
-!- SUBUNIT: Part of the E2F6.com-1 complex in G0 phase composed of
E2F6, MGA, MAX, TFDPI, CBX3, BAT8, EUMTASE1, RING1, RNF2, MBLR,
BAT8 and YAP2.
-!- SUBCELLULAR LOCATION: Nuclear (Probable).
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=3;
Name=1; Synonyms=A;
IsoId=Q969R5-1; Sequences=Displayed;
Name=2; Synonyms=B;
IsoId=Q969R5-2; Sequence=VSP_003904; VSP_003905;
Name=3;
IsoId=Q969R5-3; Sequence=VSP_003906; VSP_003907;
-!- SIMILARITY: Contains 1 C4-type zinc finger.
-!- SIMILARITY: Contains 4 MBL domains.
-!- CAUTION: Ref.3 sequences differ from that shown in that they seem
to include intronic sequence.
-!- CAUTION: Ref.4 sequence differs from that shown due to erroneous
gene model prediction.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; AJ305226; CAC37794.1; -;
EMBL; AJ305227; CAC37795.1; -;
EMBL; AL136564; CAB66499.1; -;
EMBL; AK074091; BAB84917.1; ALT SEQ.
EMBL; AK097052; BAC04936.1; ALT SEQ.
EMBL; AL035658; -; NOT ANNOTATED CDS.
EMBL; AL035681; CAB63071.1; ALT SEQ.
EMBL; BC017191; AAH17191.1; -;
EMBL; HGNC:18594; L3MBTL2.
DR GK; Q969R5; -;
DR InterPro; IPR004032; Mbt.
DR Pfam; PF02820; mbt; 4.
DR SMART; SM00561; Mbt; 4.
KW Transcription regulation; Chromatin regulator; Nuclear protein;
KW Zinc-finger; Metal-binding; Repeat; Alternative splicing;
KW Polymorphism.
FT ZN_FING 90 114 C4-TYPE.
FT DOMAIN 214 291 MET 1.
FT DOMAIN 327 399 MET 2.
FT DOMAIN 432 509 MET 3.
FT DOMAIN 540 611 MET 4.
FT DOMAIN 17 20 POLY-LYS.
FT DOMAIN 620 624 POLY-LYS.
FT VARSPLIC 608 614 EPATLK -> GKLPRL (in isoform 2).
FT VARSPLIC 615 705 Missing (in isoform 2).
FT VARSPLIC 608 617 EPATLKAKE -> GVGSRGPKRL (in isoform 3).
FT VARSPLIC 618 705 Missing (in isoform 3).
FT VARSPLIC 300 300 R -> W (IN dbSNP:2277846).
FT VARIANT 300 300 /FTID=VAR_015093.
FT SEQUENCE 705 AA; 79110 MW; 8FC86A440982FFA7 CRC64;
Query Match 100.0%; Score 2683; DB 1; Length 705;
Best Local Similarity 100.0%; Pred. No. 3.8e-216;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKGKVEVLNSDAVLPSPRYTWIASVQTAGYRVLLRYEGFENDASHDFWCVGVDPVHPI 50
DB 211 MKGKVEVLNSDAVLPSPRYTWIASVQTAGYRVLLRYEGFENDASHDFWCVGVDPVHPI 270
QY 61 GWCAINSKILVPRTTHAKFTDVKYLMKRLVGSRTLPVDFTKMWESMKY2PFRQMRLE 120

Db 271 GWCAINSKILVPPRTTHAKFTDKGKMLKRLVGSRTLPVDFHKKVMSMKYPPRQGMLE 330
 QY 121 VVDKQSVSTRMAVVDVTIGGRLLRLLYEDGSDDDFWCHMSPLTHPVGWSRRVGHGIM 180
 Db 331 VVDKQSVSTRMAVVDVTIGGRLLRLLYEDGSDDDFWCHMSPLTHPVGWSRRVGHGIM 390
 QY 181 SRRSDMAHPTFRKIYCDVAVPYLKKVRAVYTEGGWFEFGMKLEAIDPLNIGNICVATV 240
 Db 391 SRRSDMAHPTFRKIYCDVAVPYLKKVRAVYTEGGWFEFGMKLEAIDPLNIGNICVATV 450
 QY 241 CKVLLDGYLMI CVDGSPSTGDLWFCYHASSHAIPATFCQNDIELTPPKGYEAQTFNW 300
 Db 451 CKVLLDGYLMI CVDGSPSTGDLWFCYHASSHAIPATFCQNDIELTPPKGYEAQTFNW 510
 QY 301 ENYLEKTSKAAPSRLFNMDPCNHGFKVGMKLEAVDLMEPRLICVATKRVVHRLLSIH 360
 Db 511 ENYLEKTSKAAPSRLFNMDPCNHGFKVGMKLEAVDLMEPRLICVATKRVVHRLLSIH 570
 QY 361 DGDWSEYDQWDCESPDIYPVGCWCELTGYQLPPVAAPATPLKAKEATKKKKQFGKKR 420
 Db 571 DGDWSEYDQWDCESPDIYPVGCWCELTGYQLPPVAAPATPLKAKEATKKKKQFGKKR 630
 QY 421 KRIPPTKTRPLRQGGKPLLEDPOGARKISSEPVGGEIIIAVRVKEEHLDVASPDKASSP 480
 Db 631 KRIPPTKTRPLRQGGKPLLEDPOGARKISSEPVGGEIIIAVRVKEEHLDVASPDKASSP 690
 QY 481 ELPVSVENIKQETDD 495
 Db 691 ELPVSVENIKQETDD 705

RESULT 2

LML2_MOUSE
 ID LML2_MOUSE STANDARD; PRT; 703 AA.
 AC P59178;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Lethal(3)malignant brain tumor-like 2 protein (L(3)mbt-like 2 protein)
 DE (H-1(3)mbt-like protein).
 GN L3MBTL2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RP [1]
 SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Liver;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loguercio N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madao A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: Putative Polycomb group (PcG) protein. PcG proteins
 CC maintain the transcriptionally repressive state of genes, probably
 CC via a modification of chromatin, rendering it heritably changed in

CC its expressibility. Its association with a chromatin remodeling
 CC complex suggests that it may contribute to prevent expression of
 CC genes that trigger the cell into mitosis (By similarity).
 CC -!- SUBUNIT: Part of the E2F6-com-1 complex in G0 phase composed of
 CC E2F6, MGA, MAX, TRDP1, CBX3, BAT8, E2F6, RING1, RNF2, MBLR,
 CC BAT8 and YAF2 (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- SIMILARITY: Contains 1 C4-type zinc finger.
 CC -!- SIMILARITY: Contains 4 MBT domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: BC030864; AAH30864.1; -;
 CC MGD: MGI:2443584; 4732493N06Rik.
 CC SMART: SM00561; MBT; 4.
 CC KW Transcription regulation; Chromatin regulator; Nuclear protein;
 CC Zinc-finger; Metal-binding; Repeat.
 CC FT ZN FING 90 114
 CC FT DOMAIN 214 291 MBT 1.
 CC FT DOMAIN 327 399 MBT 2.
 CC FT DOMAIN 432 509 MBT 3.
 CC FT DOMAIN 540 611 MBT 4.
 CC FT DOMAIN 17 21 POLY-GLU.
 CC FT DOMAIN 620 624 POLY-LYS.
 CC SQ SEQUENCE 703 AA; 78954 MW; 2E9B5DF5BFDAAB9 CRC64;

Query Match 90.8%; Score 2436; DB 1; Length 703;
 Best Local Similarity 89.7%; Pred. No. 1.6e-195;
 Matches 444; Conservative 23; Mismatches 26; Indels 2; Gaps 1;
 QY 1 MKGMKVEVLNSDAVLPSRVYVVIQTAGYVLLRYEGFENDASHDFWCNLTGTVDVHPI 60
 Db 211 MKGMKVEVLNSDAVLPSRVYVVIQTAGYVLLRYEGFENDASHDFWCNLTGTVDVHPI 270
 QY 61 GWCAINSKILVPPRTTHAKFTDKGKMLKRLVGSRTLPVDFHKKVMSMKYPPRQGMLE 120
 Db 271 GWCAINSKILVPPRTTHAKFTDKGKMLKRLVGSRTLPVDFHKKVMSMKYPPRQGMLE 330
 QY 121 VVDKQSVSTRMAVVDVTIGGRLLRLLYEDGSDDDFWCHMSPLTHPVGWSRRVGHGIM 180
 Db 331 VVDKQSVSTRMAVVDVTIGGRLLRLLYEDGSDDDFWCHMSPLTHPVGWSRRVGHGIM 390
 QY 181 SRRSDMAHPTFRKIYCDVAVPYLKKVRAVYTEGGWFEFGMKLEAIDPLNIGNICVATV 240
 Db 391 SRRSDMAHPTFRKIYCDVAVPYLKKVRAVYTEGGWFEFGMKLEAIDPLNIGNICVATV 450
 QY 241 CKVLLDGYLMI CVDGSPSTGDLWFCYHASSHAIPATFCQNDIELTPPKGYEAQTFNW 300
 Db 451 CKVLLDGYLMI CVDGSPSTGDLWFCYHASSHAIPATFCQNDIELTPPKGYEAQTFNW 510
 QY 301 ENYLEKTSKAAPSRLFNMDPCNHGFKVGMKLEAVDLMEPRLICVATKRVVHRLLSIH 360
 Db 511 ENYLEKTSKAAPSRLFNMDPCNHGFKVGMKLEAVDLMEPRLICVATKRVVHRLLSIH 570
 QY 361 DGDWSEYDQWDCESPDIYPVGCWCELTGYQLPPVAAPATPLKAKEATKKKKQFGKKR 420
 Db 571 DGDWSEYDQWDCESPDIYPVGCWCELTGYQLPPVAAPATPLKAKEATKKKKQFGKKR 630
 QY 421 KRIPPTKTRPLRQGGKPLLEDPOGARKISSEPVGGEIIIAVRVKEEHLDVASPDKASSP 480
 Db 631 KRIPPTKTRPLRQGGKPLLEDPOGARKISSEPVGGEIIIAVRVKEEHLDVASPDKASSP 688
 QY 481 ELPVSVENIKQETDD 495
 Db 689 QLPPLPESIKQBRNN 703

```

RESULT 3
ID LMBT HUMAN STANDARD; PRT; 772 AA.
AC Q9Y468; Q9H1B5; Q9UG05; Q9UJB9; Q9Y4C9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lethal(3)malignant brain tumor-like protein (L(3)mbl-like) (L(3)mbl
DE protein homolog) (H-I(3)mbl protein) (H-L(3)Mbl).
GN L3MBTL OR L3MBT OR KIA00681.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Brain;
RX MEDLINE=99373015; PubMed=10445843;
RA Koga H., Mateui S.-I., Hirota T., Takebayashi S.-I., Okumura K.,
RA Sava H.;
RT "A human homolog of Drosophila lethal(3)malignant brain tumor
RT (L(3)mbl) protein associates with condensed mitotic chromosomes.";
RL Oncogene 18:3799-3809(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Uterus;
RX Koehrer K., Beyer A., Mewes H.-W., Cassenhuber J., Wiemann S.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burdill W.D., Butler A.B., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gillwilliam R., Hall R.E.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levaeslaimo M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilmings L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RN Nature 414:865-871(2001).
RN [4]
RP SEQUENCE OF 215-772 FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:159-176(1998).
CC -1- FUNCTION: Polycomb group (PcG) protein. PcG proteins maintain the
CC transcriptionally repressive state of genes, probably via a
CC modification of chromatin, rendering it heritably changed in its
CC expressibility. Probably plays a role in cell proliferation.
CC Overexpression induces multinucleated cells, suggesting that it is

```

```

required to accomplish normal mitosis.
-1- SUBCELLULAR LOCATION: Nuclear; excluded from the nucleolus. Does
not colocalizes with the PcG protein BMI1, suggesting that these
two proteins do not belong to the same complex.
-1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=4;
Name=1; Synonyms=mbt-1;
IsoId=Q9Y468-1; Sequence=Displayed;
Name=2; Synonyms=mbt-11;
IsoId=Q9Y468-2; Sequence=VSP_003902;
Name=3;
IsoId=Q9Y468-3; Sequence=VSP_003901; VSP_003902;
Name=4;
IsoId=Q9Y468-4; Sequence=VSP_003903;
-1- TISSUE SPECIFICITY: Widely expressed. Expression is reduced in
colorectal cancer cell line SW480 and promyelocytic leukemia cell
line HL-60.
-1- DEVELOPMENTAL STAGE: In interphase cells, it is scattered
throughout the nucleoplasm. In mitotic cells, it strongly
associates with condensed chromosomes from the prophase to
telophase.
-1- SIMILARITY: Contains 3 mbt domains.
-1- CAUTION: Ref.3 (CAC18508) sequences differ from that shown due to
erroneous gene model prediction.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U89358; AAC69438.1; --
EMBL; AL10279; CAB3714.1; --
EMBL; Z98752; CAC16799.1; --
EMBL; Z98752; CAC16800.1; --
EMBL; Z98752; CAC18508.1; --
EMBL; AL031681; CAC17518.1; ALT_SEQ.
EMBL; AL031681; CAB43959.1; --
EMBL; AB014581; BAA31656.1; --
PIR; T14794; T14794.
Gene; HGNC:15905; L3MBTL.
GK; Q9Y468; --
InterPro; IPR004092; Mbt.
InterPro; IPR002515; Znf_C2HC.
Pfam; PF02820; mbt; 3.
Pfam; PF01530; ZF-C2HC; 1.
SMART; SM00561; Mbt; 3.
Transcription regulation; Chromatin regulator; Zinc-finger;
DNA-binding; Nuclear protein; Repeat; Alternative splicing.
DOMAIN 242 315 Mbt 1.
DOMAIN 349 422 Mbt 2.
DOMAIN 453 526 Mbt 3.
ZN_FING 552 578 C2HC-TYPE.
VARSPLIC 1 348 Missing (in isoform 3).
VARSPLIC 709 772 /FTid=VSP_003901.
ARIVRVTHVSGKTLVWTVAQLGDLVCSDLHQBKGILETGV
HSLICSLPHTLLAKLSPASDSQY -> VRCKRVGDRAGVT
VLKTAGRCPPQRHFC (in isoform 2 and
isoform 3).
VARSPLIC 709 772 /FTid=VSP_003902.
ARIVRVTHVSGKTLVWTVAQLGDLVCSDLHQBKGILETGV
HSLICSLPHTLLAKLSPASDSQY -> MIDGEAPLLLTQAD
IVKIMSVKLGFLAKIYNAILMFKNADDTLK (in
isoform 4).
/FTid=VSP_003903.
P -> L (IN REF. 1).
LR -> MC (IN REF. 1).
L -> M (IN REF. 1).
S -> P (IN REF. 1).
SEQUENCE 772 AA; 85916 MW; 117B03A628826B29 CRC64;

```

Query Match		19.0%; Score 510; DB 1; Length 772;
Best Local Similarity		33.7%; Pred. No. 1.2e-34;
Matches 136; Conservative		53; Mismatches 168; Indels 47; Gaps 14;
QY	83	WGYLMKRLVGRSLTPVDF--HIKVESMKYFFRQGMRLVVDKQSVSRTRMAVVDTVIG 140
Db	208	WESYLEEQ--KAITAPVSLFQDSQAVTHNKGKGLMKLEGIDPQHPMSYFILTVAEYCG 265
QY	141	GLRLLYEDGDDDFWCHWMSPLHPVGSRRVCHGIMSE--RRSDMAHHPTFRKIYC 198
Db	266	YLRHLHFGYSGCHDFWVANSFDPHPAGWFKTKGKLQPKGYKEEFWSQYLRSRA 325
QY	199	DAVP--YLFKVRVAVTEGGWFEKMLRAIDPLNIGNICVATVCKVLDGVLICVDGCP 257
Db	326	QAAPKHLFVSQSHSPPLG-FQVGMKLEAVDEMNSLVCVASVDY-VDSFLVHFNDW 383
QY	258	STDGLDWFYHASSHAIPATPCQKNDIELTPPKGY-EAQTNNMNYLEKTKSAAPSRL 316
Db	384	DT--YDYWC-DFSSPIIHFVGMCCQKGLTPPDYDPDNFCWEKYLEETGASAVPTWA 440
QY	317	FNMDCPNHGFVKGMKLEAVDLMEPLICVATVKKRVVHLLSIHFDGWDSEYDQVDCSP 376
Db	441	FKVR--PPHSFLVMKLEAVDRNPALIRVASVEDVEDHRIKIHFDGWSHGDFWIDADHP 499
QY	377	DIYPVGCWELTGY-QPPVA-AEPAT-----PLKAXEATKXKKQKFKRKP----- 424
Db	500	DIHPAGWCKTGHPLOPLGPREPSSAPGCGPPLSYRSLPHTRTSKYSEHHRKCPGPG 559
QY	425	-----PKTRLRQSKKPLEDDPQARK 449
Db	560	DOSGHVTKFTAHNCLSGCPAER--NOSRLKAELSDSEASARK 601
RESULT 4		
ID	SSPO BOVIN	STANDARD; PRT: 867 AA.
AC	P98167;	
DI	01-OCT-1996 (Rel. 34, Created)	
DT	01-OCT-1996 (Rel. 34, Last sequence update)	
DT	28-FEB-2003 (Rel. 41, Last annotation update)	
DE	SCO-spondin (Fragment)	
OS	Bos taurus (Bovine)	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;	
OC	Bovidae; Bovinae; Bos.	
OX	NCBI_TaxID=9913;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Epidermocycle;	
RX	MEDLINE=98338614; PubMed=8743952;	
RA	Gobron S., Monnerie H., Meinzel R., Creveaux I., Lehmann W.,	
RA	Lamalle D., Pastugue B., Meinzel A.;	
RT	"SCO-spondin: a new member of the thrombospondin family secreted by	
RT	the subcommissural organ is a candidate in the modulation of neuronal	
RT	aggregation.";	
RL	J. Cell Sci. 109:1053-1061(1996).	
CC	- FUNCTION: INVOLVED IN THE MODULATION OF NEURONAL AGGREGATION.	
CC	- SUBCELLULAR LOCATION: Extracellular.	
CC	- TISSUE SPECIFICITY: SUBCOMMISSURAL ORGAN.	
CC	- DEVELOPMENTAL STAGE: EMBRYO.	
CC	- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.	
CC	- SIMILARITY: Contains at least 4 TSP type-1 domains.	
CC	- SIMILARITY: Contains at least 2 EGF-like domains.	
CC	- SIMILARITY: Contains at least 1 FS/8 type C domain.	
CC	- SIMILARITY: Contains at least 3 LDL-receptor class A domains.	
CC	- SIMILARITY: Contains 1 VWFC domain.	
CC	-----PVEHADTQGHPP-----PVL----- 400	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/	

or send an email to license@isb-sib.ch.

CC	EMBL; X93922; CAA63815.1; -	
CC	HSP; P01130; IAUJ.	
DR	InterPro; IPR000421; FA58 C.	
DR	InterPro; IPR002172; LDL_Receptor_A.	
DR	InterPro; IPR002913; TIL_Cystrich.	
DR	InterPro; IPR000884; TSPI.	
DR	InterPro; IPR001007; VWFC C.	
DR	Pfam; PF00754; F5_F8_type_C; 1.	
DR	Pfam; PF00057; ldl_recept_a; 3.	
DR	Pfam; PF01826; TIL; 1.	
DR	Pfam; PF00030; TSP; 1; 4.	
DR	Pfam; PF00033; VWFC; 1.	
DR	SMART; SMC0231; FA58C; 1.	
DR	SMART; SMC0192; LDLA; 3.	
DR	SMART; SMC0209; TSP1; 4.	
DR	SMART; SMC0214; VWFC; 1.	
DR	PROSITE; PS01285; FA58C_1; 1.	
DR	PROSITE; PS01286; FA58C_2; 1.	
DR	PROSITE; PS00022; FA58C_3; 1.	
DR	PROSITE; PS01209; LDLRA_1; 3.	
DR	PROSITE; PS00068; LDLRA_2; 3.	
DR	PROSITE; PS00092; TSP1; 4.	
DR	PROSITE; PS0184; VWFC_2; 1.	
KW	Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain.	
FT	NON_TER	1
FT	DOMAIN	29 87 TSP TYPE-1 1.
FT	DOMAIN	103 142 EGF-LIKE 1.
FT	DOMAIN	143 180 EGF-LIKE 2.
FT	DOMAIN	188 244 TSP TYPE-1 2.
FT	DOMAIN	245 304 VWFC.
FT	DOMAIN	344 502 F5/8 TYPE C.
FT	DOMAIN	506 544 LDL-RECEPTOR CLASS A 1.
FT	DOMAIN	663 701 LDL-RECEPTOR CLASS A 2.
FT	DOMAIN	723 761 LDL-RECEPTOR CLASS A 3.
FT	DOMAIN	816 844 TSP TYPE-1 3.
FT	DOMAIN	861 866 TSP TYPE-1 4.
FT	DISULFID	107 122 BY SIMILARITY.
FT	DISULFID	116 127 BY SIMILARITY.
FT	DISULFID	129 141 BY SIMILARITY.
FT	DISULFID	147 166 BY SIMILARITY.
FT	DISULFID	149 169 BY SIMILARITY.
FT	DISULFID	171 179 BY SIMILARITY.
FT	DISULFID	344 502 BY SIMILARITY.
FT	DISULFID	508 520 BY SIMILARITY.
FT	DISULFID	535 533 BY SIMILARITY.
FT	DISULFID	527 542 BY SIMILARITY.
FT	DISULFID	665 677 BY SIMILARITY.
FT	DISULFID	672 690 BY SIMILARITY.
FT	DISULFID	684 699 BY SIMILARITY.
FT	DISULFID	725 737 BY SIMILARITY.
FT	DISULFID	732 750 BY SIMILARITY.
FT	DISULFID	744 759 BY SIMILARITY.
FT	CARBOHYD	88 88 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	309 309 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	409 409 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	NON_TER	867 867
SQ	SEQUENCE	867 AA; 91817 MW; 9538F2108E787B49 CRC64;
Query Match		4.5%; Score 121; DB 1; Length 867;
Best Local Similarity		20.8%; Pred. No. 0.041;
Matches 75; Conservative		38; Mismatches 130; Indels 118; Gaps 17;
QY	161	WSPLIHPV-----GWSRRVGHGKMSERRSDMAHPTFRKIYCDVAPVYLFKKVAVYTEGG 216
Db	370	WAAILRPAGAPGWS-----PVEHADTQGHPP-----PVL----- 400
QY	217	WFEQMKLEADPLNIGNICVATVCKVLDGVLICVDGSPSTGLDWFCHASSHA--- 273
Db	401	-----QDLQLQPRNLGTIIYQAGS-----SDWLQ-----SSDGLHWSYRDIGHGTQP 445
QY	274	---IFF-----ATFCQKNDIBLTFFPKGYEAQTFNWNLYENLTKSKAAPSRLFN 318

Db 446 APQLPFWKNGPSTVWMPFARMVQARHVRWPSDGH-----QAAPSSDAN 490
QY 319 MDCENHGFVGKMGKLEAVDLMSPRLICVATKRVVHLLSIHF-----DGWD 364
Db 491 LGGP-----LRVELLG-CEPAPCLG-----VGRVCSGECAPGAPCDGVEDCKDGS 538
QY 365 SEYDOWDCESPDIYFVGCETLQYLOPPVAAEPATPLKAKEATKKKKQFGK--KKR 422
Db 539 EE-----GCVTPPA-GAGRIESTAWSAPSAQPCQLPPQPSSEGLAEADHMHGPGSP 592
QY 423 LPPTXTRLRGSKPFLDDEPQARKISSEPVGEIIIVAVKHEHLDVADSPDKASSPEL 482
Db 593 VPPTGKGASLGSEB---HPSPGSGVQTVTPTSQPEAQAALRPEAAVTVLPPHMPVTPEV 649
QY 483 P 483
Db 650 P 650
RESULT 5
ID R129 MYCGE STANDARD; PRT; 200 AA.
AC P47405;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L29.
GN RPMC OR RPL29 OR MG159.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96023346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bait C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
CC -!- SIMILARITY: BELONGS TO THE L29P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U39696; AAC71377.1; -
CC FIR; F64217; F64217.
CC TIGR; MG159; -
CC HAMAP; MF_00374; fused; 1.
DR InterPro; IPR001854; Ribosomal L29.
DR Pfam; PF00831; Ribosomal L29; 1.
DR TIGRFAMs; TIGR00012; L29; 1.
DR PROSITE; PS00579; RIBOSOMAL_L29; 1.
KW Ribosomal protein. Complete proteome.
FT DOMAIN 1 77 50S RIBOSOMAL PROTEIN L29.
FT DOMAIN 78 200 UNKNOWN.
SQ SEQUENCE 200 AA; 23258 MW; 5BC115E52AA5EA92 CRC64;
Query Match 4.3%; Score 116.5; DB 1; Length 200;
Best Local Similarity 25.0%; Pred. No. 0.04; Indels 29; Gaps 6;
Matches 44; Conservative 38; Mismatches 65;
QY 331 KLEAVDLMEPRILICVATKRVVHLLSIHFD---QWDSEYDQW--VDCESPIYFVGW-- 383

Db 32 KLAHGELDKPHLI--AKVRKLLAVVLTILTERKLNQVEKDKYKLLSRKTNELIVNSMKQ 89
QY 384 -----CELTGYLOPPVAAEPATPLKAKEATKKKKQFGKRRIPPTKTRPLRQG 434
Db 90 KUSTPESKQETKKAEPKVESKESKQETKKAEPKELKQETKKEVVKVPKVEPLKQE 149
QY 435 SKKPLEDDPQARKISSEPVGEIIIVAVKHEHLDVADSPDKASSPELPVSVENIK 490
Db 150 TKK--VEARIETKTKVESKPLKQEVKVEAKK-----SVSKPQKVPVKARMIK 194
RESULT 6
TACC2_HUMAN
ID TACC2_HUMAN STANDARD; PRT; 1026 AA.
AC O95359; Q9NZ41; Q9NZRS;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transforming acidic coiled-coil-containing protein 2 (Anti Zuai-1)
DE (AZU-1).
GN TACC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain, Fetal brain, and skeletal muscle;
RX MEDLINE=20570483; PubMed=11121038;
RA Gergely F., Karlsson C., Still I.H., Cowell J.K., Kilmartin J.,
RA Raff J.W.;
RT "The TACC domain identifies a family of centrosomal proteins that can
RT interact with microtubules."
RL Proc. Natl. Acad. Sci. U.S.A. 97:14352-14357(2000).
RN [2]
RP SEQUENCE OF 375-1026 FROM N.A. (ISOFORM 2).
RX MEDLINE=20214926; PubMed=1074935;
RA Chen H.-M., Schmeichel K.L., Mian I.S., Lelievre S., Petersen O.W.,
RA Bissell M.J.;
RT "AZU-1: a candidate breast tumor suppressor and biomarker for tumor
RT progression."
RL Mol. Biol. Cell 11:1357-1367(2000).
RN [3]
RP SEQUENCE OF 44-1026 FROM N.A. (ISOFORM 3).
RA Pu J., Li C., Rodriguez M., Banerjee D.;
RT "Expression of TACC2 protein mRNA in human microvascular endothelial
RT cells."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY PLAY A ROLE IN ORGANIZING CENTROSOMAL MICROTUBULES.
CC MAY ACT AS A TUMOR SUPPRESSOR PROTEIN. MAY REPRESENT A TUMOR
CC PROGRESSION MARKER.
CC -!- SUBUNIT: INTERACTS WITH MICROTUBULES.
CC -!- SUBCELLULAR LOCATION: NUCLEAR; CONCENTRATED AT CENTROSOMES.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Experimental confirmation may be lacking for some
CC isoforms;
CC Name=1;
CC IsoId=O95359-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O95359-2; Sequence=VSP_006368, VSP_006369;
CC Name=3;
CC IsoId=O95359-3; Sequence=VSP_006369;
CC -!- SIMILARITY: BELONGS TO THE TACC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC


```
DR EMBL; AF095791; AAC64968.2; --
DR EMBL; AF176646; AAF63433.1; ALT_INIT.
DR EMBL; AF220152; AAF29537.2; --
DR Genew; HGNC:11523; TACC2.
DR MIM; 605302; --
DR GO; GO:0007048; P:oncogenesis; TAS.
DR Pfam; PF05010; TACC; 1.
KW Coiled coil; Nuclear protein; Alternative splicing.
FT DOMAIN 43 51 POLY-PRO.
FT DOMAIN 498 501 POLY-LYS.
FT DOMAIN 753 781 COILED COIL (POTENTIAL).
FT DOMAIN 824 1025 COILED COIL (POTENTIAL).
FT VARSPPLIC 507 510 Missing (in isoform 2).
FT VARSPPLIC 507 510 Missing (in isoform 2).
FT VARSPPLIC 711 787 Missing (in isoform 2 and isoform 3).
FT VARSPPLIC 711 787 /FTID-VSP 006368.
FT CONFLICT 44 67 /FTID-VSP 006369.
FT CONFLICT 44 67 PAPPPPPPEV;PEVSTQPPPEE -> MFYKRRGAPMRP
FT VSVTDGVVCS (IN REF. 3).
SQ SEQUENCE 1026 AA; 112110 MW; E2575FCB446E9CF8 CRC64;

Query Match 4.1%; Score 109.5; DB 1; Length 1026;
Best Local Similarity 23.6%; Pred. No. 0.47;
Matches 57; Conservative 33; Mismatches 95; Indels 57; Gaps 11;

QY 253 VDGSPSTGLDFWFCVHASSHALFATFCQKNDIELTPKGYEAOTFNWENYLEKTKSKAA 312
DQ 77 VPDGRSDVSGSPFRPSSH--FSAVFDEKPIASSGTYNLDFNIELVDTFTQLEPRAS 135
QY 313 PSRLFNMDPNHGFKVGMKLEAVLMEPRLLICVATKRVVHRLLSIH---FDGWDSDYDQ 369
DQ 136 -----DAKQEGKVNRKSTD-----SVISKSTLSLSLQASDFDQASSSGNP 181
QY 370 WDCESPIYPVWCELTGYQLPPVAAPATPKAKEATKK-----KKQFGKKKKRI 423
DQ 182 EAVLAPAYATG-----SSASSTLK--RTKKRPPSLKKQTTK--- 221
QY 424 PPTKTRPQRSKKPLLEDPPQGARKISSEVPV-GEIIAVRVKHEHLDVASPKASSPEL 482
DQ 222 -PTETPPVKETQOE-----DRESL-----VPSGENLASETKTESAKTEGSPALLEET 269
QY 483 PV 484
DQ 270 PL 271

RESULT 7
NEBU_HUMAN STANDARD; PRT; 6669 AA.
AC P20929; Q15346;
DT 01-FEB-1991 (Rel. 17, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nebulin.
GN NEB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. PubMed=7739042;
RX MEDLINE=95257391; PubMed=7739042;
RA Label S.; Kolmerer B.;
RT "The complete primary structure of human nebulin and its correlation
to muscle structure.";
RL J. Mol. Biol. 248:308-315 (1995).
RN [2]
RP PARTIAL PRELIMINARY SEQUENCE.
RX MEDLINE=88284704; PubMed=3397062;
RA Zeviani M., Darras B.T., Rizzuto R., Salviati G., Betto R.,
RA Bonilla E., Miranda A.F., Du J., Samitt C., Dickson G., Walsh F.S.,
RA Dimauro S., Francke U., Schon E.A.;
RT "Cloning and expression of human nebulin cDNAs and assignment of the
gene to chromosome 2q31-q32.";
```

```
Genomics 2:249-256 (1988).
[3]
STRUCTURE BY NMR OF 6610-6669.
MEDLINE=98179559; PubMed=9514727;
RA Politou A.S.; Millevoi S.; Gautel M.; Kolmerer B.; Pastore A.;
RT "SH3 in muscles: solution structure of the SH3 domain from nebulin.";
J. Mol. Biol. 276:189-202 (1998).
CC -1- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MAINTAINING
THE STRUCTURAL INTEGRITY OF SARCOMERES AND THE MEMBRANE SYSTEM
ASSOCIATED WITH THE MYOFIBRILS. BIND AND STABILIZE F-ACTIN.
CC -1- TISSUE SPECIFICITY: MUSCLE-SPECIFIC. LOCATED IN THE THIN FILAMENT
OF STRIATED MUSCLE.
CC -1- DISEASE: Defects in NEB are a cause of the autosomal recessive
form of nemaline myopathy (NEM2).
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -1- SIMILARITY: Contains 178 nebulin repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X83957; CAAS8788.1; --
EMBL; M19668; AAAS9916.1; ALT_SEQ.
EMBL; M19669; AAAS9917.1; ALT_SEQ.
DR PIR; S55024; S55024.
DR PDB; 1ARK; 28-JAN-98.
DR PDB; 1NEB; 24-DEC-97.
DR Genew; HGNC:7720; NEB.
DR MIM; 161650; --
DR GO; GO:0015629; C:actin cytoskeleton; TAS.
DR GO; GO:0030017; C:sarcomere; NAS.
DR GO; GO:0003792; F:regulation of actin thin filament length ac. .; NAS.
DR GO; GO:0006307; F:structural constituent of muscle; TAS.
DR GO; GO:0007525; P:somatic muscle development; NAS.
DR InterPro; IPR000900; Nebulin.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00880; Nebulin; 146.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PRO0510; NEBULIN.
DR PRINTS; PRO0452; SH3DOMAIN.
DR ProDom; PD0000666; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KW Actin-binding; Muscle protein; Cytoskeleton; Repeat; 3D-structure.
FT REPEAT 76 107 NEBULIN 1.
FT REPEAT 113 143 NEBULIN 2.
FT REPEAT 148 178 NEBULIN 3.
FT REPEAT 183 213 NEBULIN 4.
FT REPEAT 218 248 NEBULIN 5.
FT REPEAT 253 283 NEBULIN 6.
FT REPEAT 289 318 NEBULIN 7.
FT REPEAT 324 354 NEBULIN 8.
FT REPEAT 363 393 NEBULIN 9.
FT REPEAT 398 428 NEBULIN 10.
FT REPEAT 434 464 NEBULIN 11.
FT REPEAT 502 532 NEBULIN 12.
FT REPEAT 537 567 NEBULIN 13.
FT REPEAT 573 603 NEBULIN 14.
FT REPEAT 611 641 NEBULIN 15.
FT REPEAT 681 711 NEBULIN 16.
FT REPEAT 749 779 NEBULIN 17.
FT REPEAT 784 814 NEBULIN 18.
FT REPEAT 820 850 NEBULIN 19.
FT REPEAT 858 888 NEBULIN 20.
FT REPEAT 893 923 NEBULIN 21.
FT REPEAT 924 954 NEBULIN 22.
FT REPEAT 959 990 NEBULIN 23.
FT REPEAT 993 1023 NEBULIN 24.
FT REPEAT 1028 1058 NEBULIN 25.
```


CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Bovinae; Bos.
 CC NCBI_TaxID=9913;
 RN [1] SEQUENCE FROM N.A., AND SEQUENCE OF 1-20; 206-254 AND 385-396.
 RP TISSUE=thymus;
 RC MEDLINE=92097545; PubMed=1756732;
 RA Wahle E., Martin G., Schiltz E., Keller W.;
 RT "Isolation and expression of cDNA clones encoding mammalian poly(A)
 RT polymerase.";
 RL EMBO J. 10:4251-4257(1991).
 RN [2] REVISIONS.
 RP REVISIONS.
 RA Wahle E.;
 RL Submitted (AUG-1992) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 96-102 AND 444-467.
 RC TISSUE=Heart muscle;
 EX MEDLINE=91375535; PubMed=1896071;
 RA Raabe T., Bollum F.J., Manley J.L.;
 RT "Primary structure and expression of bovine poly(A) polymerase.";
 RL Nature 353:229-234(1991).
 RN [4]
 RP DOMAINS, ACTIVE SITES, AND MUTAGENESIS.
 EX MEDLINE=96221320; PubMed=8665867;
 RA Martin G., Keller W.;
 RT "Mutation analysis of mammalian poly(A) polymerase identifies a
 RT region for primer binding and catalytic domain, homologous to the
 RT family X polymerases, and to other nucleotidyltransferases.";
 RL EMBO J. 15:2593-2603(1996).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 EX MEDLINE=20402325; PubMed=10944102;
 RA Martin G., Keller W., Double S.;
 RT "Crystal structure of mammalian poly(A) polymerase in complex with an
 RT analog of ATP.";
 RL EMBO J. 19:4193-4203(2000).
 CC -!- FUNCTION: Polymerase that creates the 3' poly(A) tail of mRNA's.
 CC Also required for the endoribonucleolytic cleavage reaction at
 CC some polyadenylation sites. May acquire specificity through
 CC interaction with a cleavage and polyadenylation specificity factor
 CC (CPSF) at its C-terminus.
 CC -!- CATALYTIC ACTIVITY: N ATP + [nucleotide] (M) = N diphosphate +
 CC [nucleotide] (M+N).
 CC -!- SUBUNIT: Monomer.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=Long;
 CC IsoId=P25500-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=P25500-2; Sequence=VSP_004524, VSP_004525, VSP_004526;
 CC Note=No experimental confirmation available;
 CC -!- PTM: Phosphorylated. Phosphorylation/dephosphorylation may
 CC regulate the interaction between PAP and CPSF.
 CC -!- SIMILARITY: BELONGS TO THE POLY(A) POLYMERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X61585; CAA43782.1; -;
 CC EMBL; X63436; CAA45031.1; -;
 CC PDB; 1F5A; 13-SEP-00.
 DR InterPro; IPR002934; NTP transf.
 DR InterPro; IPR001201; PAP_25A core.
 DR InterPro; IPR007012; PAP_cent.

DR InterPro; IPR007010; PAP RNA bind.
 DR Pfam; PF01909; NTP transf. 2; 1.
 DR Pfam; PF04928; PAP_central; 1.
 DR Pfam; PF04926; PAP_RNA-bind; 1.
 KW mRNA processing; Transference; Transcription; RNA-binding;
 KW Nuclear protein; Phosphorylation; Alternative splicing; 3D-structure.
 FT INIT_MET 0 0
 FT ACT_SITE 112 112
 FT ACT_SITE 114 114
 FT ACT_SITE 166 166
 FT DOMAIN 499 506
 FT DOMAIN 643 658
 FT VARSPIC 662 682
 FT VARSPIC 708 709
 FT VARSPIC 710 738
 FT CONFLICT 79 79
 FT HELIX 32 45
 FT HELIX 46 48
 FT TURN 49 49
 FT HELIX 81 81
 FT TURN 82 83
 FT HELIX 86 89
 FT TURN 90 91
 FT STRAND 95 98
 FT HELIX 100 104
 FT TURN 105 105
 FT TURN 109 110
 FT STRAND 113 119
 FT TURN 121 122
 FT HELIX 125 128
 FT TURN 129 130
 FT HELIX 131 137
 FT TURN 138 138
 FT TURN 140 141
 FT STRAND 142 148
 FT TURN 149 150
 FT STRAND 155 160
 FT TURN 161 162
 FT STRAND 163 171
 FT TURN 179 180
 FT TURN 183 184
 FT HELIX 186 188
 FT TURN 189 191
 FT HELIX 194 211
 FT TURN 212 212
 FT HELIX 216 232
 FT TURN 233 234
 FT STRAND 236 236
 FT TURN 238 241
 FT STRAND 243 243
 FT HELIX 245 258
 FT TURN 260 261
 FT HELIX 264 276
 FT TURN 277 277
 FT TURN 280 281
 FT STRAND 284 284
 FT TURN 301 303
 FT HELIX 305 309
 FT STRAND 314 315
 FT STRAND 323 323
 FT TURN 324 327
 FT HELIX 330 351
 FT TURN 352 353
 FT HELIX 357 359
 FT TURN 360 361
 FT HELIX 366 369
 FT STRAND 372 379
 FT HELIX 383 394
 FT TURN 395 396
 FT HELIX 397 405

InterPro; IPR007010; PAP RNA bind.
 Pfam; PF01909; NTP transf. 2; 1.
 Pfam; PF04928; PAP_central; 1.
 Pfam; PF04926; PAP_RNA-bind; 1.
 mRNA processing; Transference; Transcription; RNA-binding;
 Nuclear protein; Phosphorylation; Alternative splicing; 3D-structure.
 INIT_MET 0 0
 ACT_SITE 112 112
 ACT_SITE 114 114
 ACT_SITE 166 166
 DOMAIN 499 506
 DOMAIN 643 658
 VARSPIC 662 682
 VARSPIC 708 709
 VARSPIC 710 738
 CONFLICT 79 79
 HELIX 32 45
 HELIX 46 48
 TURN 49 49
 HELIX 81 81
 TURN 82 83
 HELIX 86 89
 TURN 90 91
 STRAND 95 98
 HELIX 100 104
 TURN 105 105
 TURN 109 110
 STRAND 113 119
 TURN 121 122
 HELIX 125 128
 TURN 129 130
 HELIX 131 137
 TURN 138 138
 TURN 140 141
 STRAND 142 148
 TURN 149 150
 STRAND 155 160
 TURN 161 162
 STRAND 163 171
 TURN 179 180
 TURN 183 184
 HELIX 186 188
 TURN 189 191
 HELIX 194 211
 TURN 212 212
 HELIX 216 232
 TURN 233 234
 STRAND 236 236
 TURN 238 241
 STRAND 243 243
 HELIX 245 258
 TURN 260 261
 HELIX 264 276
 TURN 277 277
 TURN 280 281
 STRAND 284 284
 TURN 301 303
 HELIX 305 309
 STRAND 314 315
 STRAND 323 323
 TURN 324 327
 HELIX 330 351
 TURN 352 353
 HELIX 357 359
 TURN 360 361
 HELIX 366 369
 STRAND 372 379
 HELIX 383 394
 TURN 395 396
 HELIX 397 405

NUCLEAR LOCALIZATION SIGNAL 1.
 NUCLEAR LOCALIZATION SIGNAL 2.
 Missing (in isoform Short).
 /FTId=VSP_004524.
 KT -> II (in isoform Short).
 /FTId=VSP_004525.
 Missing (in isoform Short).
 /FTId=VSP_004526.
 S -> R (IN REF. 3).

```
FT TURN 406 406
FT TURN 408 409
FT STRAND 410 415
FT STRAND 420 421
FT STRAND 433 442
FT HELIX 456 472
FT TURN 473 474
FT TURN 478 479
FT STRAND 481 488
FT HELIX 489 491
FT HELIX 493 495
SQ SEQUENCE 738 AA; 82310 MW; DOB90662F89363E0 CRC64;

Query Match 3.7%; Score 98.5; DB 1; Length 738;
Best Local Similarity 20.9%; Fred. No. 2.5;
Matches 53; Conservative 33; Mismatches 105; Indels 63; Gaps 9;

Qy 274 IFPATFCQKN-DIELTPPKGYEAQTFNWNYLEKTSKAAPSLFNMDCPNHGFKVGMKL 332
Db 441 VFEXTENSENLSDLT-----YDIQSFTDVTYRQAINSKM-----FEYDMKI 482

Qy 333 EAVDLMEPRILICVATKRVVHRLLSIH-----DGW-----DSYDQWDCESP 377
Db 483 AAMHVKKKQL-----HQLPSHVLOKKKHSTGVKLTPLNDSSLDLSMDSNDNM 532

Qy 378 IYPVGWCELTCYOLQ-----PPVAARPATPLKAKBAT-----KKKKQFGKKK 421
Db 533 SVSPSTAMKTSPLNSGSGQGNPAPAVTAASVTNIQTVEVSLPQINSSESSGGTSS 592

Qy 422 RIPTKTRPRQSKKFLLEDDPQGARKISSEVPVGEIIAVRVKEHLVDVAPDKASSPE 481
Db 593 SIPTATQPAISPPKPTSVRSVSTLVNPPRPSSGNAAKIPN---PIVGKRTSSPH 649

Qy 482 LPVSVENIKQETDD 495
Db 650 KEESPXKTKTEDE 663

RESULT 9
PAP_MOUSE
ID_ PAP_MOUSE STANDARD; PRT; 738 AA.
AC Q61183; Q61208; Q61209; Q8K4X2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Poly(A) polymerase alpha (EC 2.7.7.19) (PAP) (Polynucleotide
DE adenylyltransferase)
GN PAPOLA OR PAP OR PLAP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 2; 3 AND 4).
RC TISSUE=Brain;
RX MEDLINE=96189120; PubMed=8628305;
RA Zhao W., Manley J.L.;
RT "Complex alternative RNA processing generates an unexpected diversity
RL of poly(A) polymerase isoforms.";
RN [2] Mol. Cell. Biol. 16:2378-2386(1996).
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=129/Svj;
RX MEDLINE=22359775; PubMed=12471261;
RA Kashiwabara S., Noguchi J., Zhuang T., Ohmura K., Honda A., Baba T.;
RA Sugiura S., Miyamoto K., Takahashi S., Inoue K., Ogura A.;
RT "Regulation of spermatogenesis by testis-specific, cytoplasmic
RL poly(A) polymerase TPAP";
RL Science 298:1999-2002(2002).
CC -1- FUNCTION: Polymerase that creates the 3' poly(A) tail of mRNA's.
CC Also required for the endoribonucleolytic cleavage reaction at
CC some polyadenylation sites. May acquire specificity through
CC interaction with a cleavage and polyadenylation specificity factor
```

```
CC (CPSF) at its C-terminus (By similarity).
CC -1- CATALYTIC ACTIVITY: N ATP + {nucleotide}(M) = N diphosphate +
CC {nucleotide}(M+N).
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1;
CC IsoId=Q61183-1; Sequence=Displayed;
CC Name=2; Synonyms=III;
CC IsoId=Q61183-2; Sequence=VSP_004531, VSP_004532;
CC Name=3; Synonyms=V;
CC IsoId=Q61183-3; Sequence=VSP_004529, VSP_004530;
CC Name=4; Synonyms=VI;
CC IsoId=Q61183-4; Sequence=VSP_004527, VSP_004528;
CC -1- PTM: Phosphorylated. Phosphorylation/dephosphorylation may
CC regulate the interaction between PAP and CPSF (By similarity).
CC -1- SIMILARITY: BELONGS TO THE POLY(A) POLYMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U52197; AAC52586.1; -.
CC EMBL; U58134; AAC52608.1; -.
CC EMBL; U58135; AAC52609.1; -.
CC EMBL; AB086650; BAC00996.1; -.
CC EMBL; AB086633; BAC00996.1; JOINED.
CC EMBL; AB086634; BAC00996.1; JOINED.
CC EMBL; AB086635; BAC00996.1; JOINED.
CC EMBL; AB086636; BAC00996.1; JOINED.
CC EMBL; AB086637; BAC00996.1; JOINED.
CC EMBL; AB086638; BAC00996.1; JOINED.
CC EMBL; AB086639; BAC00996.1; JOINED.
CC EMBL; AB086640; BAC00996.1; JOINED.
CC EMBL; AB086641; BAC00996.1; JOINED.
CC EMBL; AB086642; BAC00996.1; JOINED.
CC EMBL; AB086643; BAC00996.1; JOINED.
CC EMBL; AB086644; BAC00996.1; JOINED.
CC EMBL; AB086645; BAC00996.1; JOINED.
CC EMBL; AB086646; BAC00996.1; JOINED.
CC EMBL; AB086647; BAC00996.1; JOINED.
CC EMBL; AB086648; BAC00996.1; JOINED.
CC EMBL; AB086649; BAC00996.1; JOINED.
CC HSP; P25500; 1F5A.
CC MGD; MG1:109301; Papola.
CC InterPro; IPR002934; NTP transf.
CC InterPro; IPR001201; PAP_25A_core.
CC InterPro; IPR007012; PAP-centr.
CC InterPro; IPR007010; PAP-RNA_bind.
CC Pfam; PF04926; PAP-RNA-bind; 1.
CC mRNA processing; Transferase; Transcription; RNA-binding;
CC Phosphorylation; Nuclear protein; Alternative splicing.
CC INIT MET 0 0 BY SIMILARITY.
CC ACT_SITE 112 112 BY SIMILARITY.
CC ACT_SITE 114 114 BY SIMILARITY.
CC ACT_SITE 166 166 BY SIMILARITY.
CC VARSPPLIC 279 292 EWPNFVLKQFEC -> YVFRLYNKKIDCRH (in
CC isoform 4).
CC /FtId=VSP_004527.
CC /FtId=VSP_004528.
CC /FtId=VSP_004529.
CC /FtId=VSP_004530.
CC /FtId=VSP_004531.
CC /FtId=VSP_004532.
CC /FtId=VSP_004533.
CC /FtId=VSP_004534.
CC /FtId=VSP_004535.
CC /FtId=VSP_004536.
CC /FtId=VSP_004537.
CC /FtId=VSP_004538.
CC /FtId=VSP_004539.
CC /FtId=VSP_004540.
CC /FtId=VSP_004541.
CC /FtId=VSP_004542.
CC /FtId=VSP_004543.
CC /FtId=VSP_004544.
CC /FtId=VSP_004545.
CC /FtId=VSP_004546.
CC /FtId=VSP_004547.
CC /FtId=VSP_004548.
CC /FtId=VSP_004549.
CC /FtId=VSP_004550.
CC /FtId=VSP_004551.
CC /FtId=VSP_004552.
CC /FtId=VSP_004553.
CC /FtId=VSP_004554.
CC /FtId=VSP_004555.
CC /FtId=VSP_004556.
CC /FtId=VSP_004557.
CC /FtId=VSP_004558.
CC /FtId=VSP_004559.
CC /FtId=VSP_004560.
CC /FtId=VSP_004561.
CC /FtId=VSP_004562.
CC /FtId=VSP_004563.
CC /FtId=VSP_004564.
CC /FtId=VSP_004565.
CC /FtId=VSP_004566.
CC /FtId=VSP_004567.
CC /FtId=VSP_004568.
CC /FtId=VSP_004569.
CC /FtId=VSP_004570.
CC /FtId=VSP_004571.
CC /FtId=VSP_004572.
CC /FtId=VSP_004573.
CC /FtId=VSP_004574.
CC /FtId=VSP_004575.
CC /FtId=VSP_004576.
CC /FtId=VSP_004577.
CC /FtId=VSP_004578.
CC /FtId=VSP_004579.
CC /FtId=VSP_004580.
CC /FtId=VSP_004581.
CC /FtId=VSP_004582.
CC /FtId=VSP_004583.
CC /FtId=VSP_004584.
CC /FtId=VSP_004585.
CC /FtId=VSP_004586.
CC /FtId=VSP_004587.
CC /FtId=VSP_004588.
CC /FtId=VSP_004589.
CC /FtId=VSP_004590.
CC /FtId=VSP_004591.
CC /FtId=VSP_004592.
CC /FtId=VSP_004593.
CC /FtId=VSP_004594.
CC /FtId=VSP_004595.
CC /FtId=VSP_004596.
CC /FtId=VSP_004597.
CC /FtId=VSP_004598.
CC /FtId=VSP_004599.
CC /FtId=VSP_004600.
CC /FtId=VSP_004601.
CC /FtId=VSP_004602.
CC /FtId=VSP_004603.
CC /FtId=VSP_004604.
CC /FtId=VSP_004605.
CC /FtId=VSP_004606.
CC /FtId=VSP_004607.
CC /FtId=VSP_004608.
CC /FtId=VSP_004609.
CC /FtId=VSP_004610.
CC /FtId=VSP_004611.
CC /FtId=VSP_004612.
CC /FtId=VSP_004613.
CC /FtId=VSP_004614.
CC /FtId=VSP_004615.
CC /FtId=VSP_004616.
CC /FtId=VSP_004617.
CC /FtId=VSP_004618.
CC /FtId=VSP_004619.
CC /FtId=VSP_004620.
CC /FtId=VSP_004621.
CC /FtId=VSP_004622.
CC /FtId=VSP_004623.
CC /FtId=VSP_004624.
CC /FtId=VSP_004625.
CC /FtId=VSP_004626.
CC /FtId=VSP_004627.
CC /FtId=VSP_004628.
CC /FtId=VSP_004629.
CC /FtId=VSP_004630.
CC /FtId=VSP_004631.
CC /FtId=VSP_004632.
CC /FtId=VSP_004633.
CC /FtId=VSP_004634.
CC /FtId=VSP_004635.
CC /FtId=VSP_004636.
CC /FtId=VSP_004637.
CC /FtId=VSP_004638.
CC /FtId=VSP_004639.
CC /FtId=VSP_004640.
CC /FtId=VSP_004641.
CC /FtId=VSP_004642.
CC /FtId=VSP_004643.
CC /FtId=VSP_004644.
CC /FtId=VSP_004645.
CC /FtId=VSP_004646.
CC /FtId=VSP_004647.
CC /FtId=VSP_004648.
CC /FtId=VSP_004649.
CC /FtId=VSP_004650.
CC /FtId=VSP_004651.
CC /FtId=VSP_004652.
CC /FtId=VSP_004653.
CC /FtId=VSP_004654.
CC /FtId=VSP_004655.
CC /FtId=VSP_004656.
CC /FtId=VSP_004657.
CC /FtId=VSP_004658.
CC /FtId=VSP_004659.
CC /FtId=VSP_004660.
CC /FtId=VSP_004661.
CC /FtId=VSP_004662.
CC /FtId=VSP_004663.
CC /FtId=VSP_004664.
CC /FtId=VSP_004665.
CC /FtId=VSP_004666.
CC /FtId=VSP_004667.
CC /FtId=VSP_004668.
CC /FtId=VSP_004669.
CC /FtId=VSP_004670.
CC /FtId=VSP_004671.
CC /FtId=VSP_004672.
CC /FtId=VSP_004673.
CC /FtId=VSP_004674.
CC /FtId=VSP_004675.
CC /FtId=VSP_004676.
CC /FtId=VSP_004677.
CC /FtId=VSP_004678.
CC /FtId=VSP_004679.
CC /FtId=VSP_004680.
CC /FtId=VSP_004681.
CC /FtId=VSP_004682.
CC /FtId=VSP_004683.
CC /FtId=VSP_004684.
CC /FtId=VSP_004685.
CC /FtId=VSP_004686.
CC /FtId=VSP_004687.
CC /FtId=VSP_004688.
CC /FtId=VSP_004689.
CC /FtId=VSP_004690.
CC /FtId=VSP_004691.
CC /FtId=VSP_004692.
CC /FtId=VSP_004693.
CC /FtId=VSP_004694.
CC /FtId=VSP_004695.
CC /FtId=VSP_004696.
CC /FtId=VSP_004697.
CC /FtId=VSP_004698.
CC /FtId=VSP_004699.
CC /FtId=VSP_004700.
CC /FtId=VSP_004701.
CC /FtId=VSP_004702.
CC /FtId=VSP_004703.
CC /FtId=VSP_004704.
CC /FtId=VSP_004705.
CC /FtId=VSP_004706.
CC /FtId=VSP_004707.
CC /FtId=VSP_004708.
CC /FtId=VSP_004709.
CC /FtId=VSP_004710.
CC /FtId=VSP_004711.
CC /FtId=VSP_004712.
CC /FtId=VSP_004713.
CC /FtId=VSP_004714.
CC /FtId=VSP_004715.
CC /FtId=VSP_004716.
CC /FtId=VSP_004717.
CC /FtId=VSP_004718.
CC /FtId=VSP_004719.
CC /FtId=VSP_004720.
CC /FtId=VSP_004721.
CC /FtId=VSP_004722.
CC /FtId=VSP_004723.
CC /FtId=VSP_004724.
CC /FtId=VSP_004725.
CC /FtId=VSP_004726.
CC /FtId=VSP_004727.
CC /FtId=VSP_004728.
CC /FtId=VSP_004729.
CC /FtId=VSP_004730.
CC /FtId=VSP_004731.
CC /FtId=VSP_004732.
CC /FtId=VSP_004733.
CC /FtId=VSP_004734.
CC /FtId=VSP_004735.
CC /FtId=VSP_004736.
CC /FtId=VSP_004737.
CC /FtId=VSP_004738.
CC /FtId=VSP_004739.
CC /FtId=VSP_004740.
CC /FtId=VSP_004741.
CC /FtId=VSP_004742.
CC /FtId=VSP_004743.
CC /FtId=VSP_004744.
CC /FtId=VSP_004745.
CC /FtId=VSP_004746.
CC /FtId=VSP_004747.
CC /FtId=VSP_004748.
CC /FtId=VSP_004749.
CC /FtId=VSP_004750.
CC /FtId=VSP_004751.
CC /FtId=VSP_004752.
CC /FtId=VSP_004753.
CC /FtId=VSP_004754.
CC /FtId=VSP_004755.
CC /FtId=VSP_004756.
CC /FtId=VSP_004757.
CC /FtId=VSP_004758.
CC /FtId=VSP_004759.
CC /FtId=VSP_004760.
CC /FtId=VSP_004761.
CC /FtId=VSP_004762.
CC /FtId=VSP_004763.
CC /FtId=VSP_004764.
CC /FtId=VSP_004765.
CC /FtId=VSP_004766.
CC /FtId=VSP_004767.
CC /FtId=VSP_004768.
CC /FtId=VSP_004769.
CC /FtId=VSP_004770.
CC /FtId=VSP_004771.
CC /FtId=VSP_004772.
CC /FtId=VSP_004773.
CC /FtId=VSP_004774.
CC /FtId=VSP_004775.
CC /FtId=VSP_004776.
CC /FtId=VSP_004777.
CC /FtId=VSP_004778.
CC /FtId=VSP_004779.
CC /FtId=VSP_004780.
CC /FtId=VSP_004781.
CC /FtId=VSP_004782.
CC /FtId=VSP_004783.
CC /FtId=VSP_004784.
CC /FtId=VSP_004785.
CC /FtId=VSP_004786.
CC /FtId=VSP_004787.
CC /FtId=VSP_004788.
CC /FtId=VSP_004789.
CC /FtId=VSP_004790.
CC /FtId=VSP_004791.
CC /FtId=VSP_004792.
CC /FtId=VSP_004793.
CC /FtId=VSP_004794.
CC /FtId=VSP_004795.
CC /FtId=VSP_004796.
CC /FtId=VSP_004797.
CC /FtId=VSP_004798.
CC /FtId=VSP_004799.
CC /FtId=VSP_004800.
CC /FtId=VSP_004801.
CC /FtId=VSP_004802.
CC /FtId=VSP_004803.
CC /FtId=VSP_004804.
CC /FtId=VSP_004805.
CC /FtId=VSP_004806.
CC /FtId=VSP_004807.
CC /FtId=VSP_004808.
CC /FtId=VSP_004809.
CC /FtId=VSP_004810.
CC /FtId=VSP_004811.
CC /FtId=VSP_004812.
CC /FtId=VSP_004813.
CC /FtId=VSP_004814.
CC /FtId=VSP_004815.
CC /FtId=VSP_004816.
CC /FtId=VSP_004817.
CC /FtId=VSP_004818.
CC /FtId=VSP_004819.
CC /FtId=VSP_004820.
CC /FtId=VSP_004821.
CC /FtId=VSP_004822.
CC /FtId=VSP_004823.
CC /FtId=VSP_004824.
CC /FtId=VSP_004825.
CC /FtId=VSP_004826.
CC /FtId=VSP_004827.
CC /FtId=VSP_004828.
CC /FtId=VSP_004829.
CC /FtId=VSP_004830.
CC /FtId=VSP_004831.
CC /FtId=VSP_004832.
CC /FtId=VSP_004833.
CC /FtId=VSP_004834.
CC /FtId=VSP_004835.
CC /FtId=VSP_004836.
CC /FtId=VSP_004837.
CC /FtId=VSP_004838.
CC /FtId=VSP_004839.
CC /FtId=VSP_004840.
CC /FtId=VSP_004841.
CC /FtId=VSP_004842.
CC /FtId=VSP_004843.
CC /FtId=VSP_004844.
CC /FtId=VSP_004845.
CC /FtId=VSP_004846.
CC /FtId=VSP_004847.
CC /FtId=VSP_004848.
CC /FtId=VSP_004849.
CC /FtId=VSP_004850.
CC /FtId=VSP_004851.
CC /FtId=VSP_004852.
CC /FtId=VSP_004853.
CC /FtId=VSP_004854.
CC /FtId=VSP_004855.
CC /FtId=VSP_004856.
CC /FtId=VSP_004857.
CC /FtId=VSP_004858.
CC /FtId=VSP_004859.
CC /FtId=VSP_004860.
CC /FtId=VSP_004861.
CC /FtId=VSP_004862.
CC /FtId=VSP_004863.
CC /FtId=VSP_004864.
CC /FtId=VSP_004865.
CC /FtId=VSP_004866.
CC /FtId=VSP_004867.
CC /FtId=VSP_004868.
CC /FtId=VSP_004869.
CC /FtId=VSP_004870.
CC /FtId=VSP_004871.
CC /FtId=VSP_004872.
CC /FtId=VSP_004873.
CC /FtId=VSP_004874.
CC /FtId=VSP_004875.
CC /FtId=VSP_004876.
CC /FtId=VSP_004877.
CC /FtId=VSP_004878.
CC /FtId=VSP_004879.
CC /FtId=VSP_004880.
CC /FtId=VSP_004881.
CC /FtId=VSP_004882.
CC /FtId=VSP_004883.
CC /FtId=VSP_004884.
CC /FtId=VSP_004885.
CC /FtId=VSP_004886.
CC /FtId=VSP_004887.
CC /FtId=VSP_004888.
CC /FtId=VSP_004889.
CC /FtId=VSP_004890.
CC /FtId=VSP_004891.
CC /FtId=VSP_004892.
CC /FtId=VSP_004893.
CC /FtId=VSP_004894.
CC /FtId=VSP_004895.
CC /FtId=VSP_004896.
CC /FtId=VSP_004897.
CC /FtId=VSP_004898.
CC /FtId=VSP_004899.
CC /FtId=VSP_004900.
CC /FtId=VSP_004901.
CC /FtId=VSP_004902.
CC /FtId=VSP_004903.
CC /FtId=VSP_004904.
CC /FtId=VSP_004905.
CC /FtId=VSP_004906.
CC /FtId=VSP_004907.
CC /FtId=VSP_004908.
CC /FtId=VSP_004909.
CC /FtId=VSP_004910.
CC /FtId=VSP_004911.
CC /FtId=VSP_004912.
CC /FtId=VSP_004913.
CC /FtId=VSP_004914.
CC /FtId=VSP_004915.
CC /FtId=VSP_004916.
CC /FtId=VSP_004917.
CC /FtId=VSP_004918.
CC /FtId=VSP_004919.
CC /FtId=VSP_004920.
CC /FtId=VSP_004921.
CC /FtId=VSP_004922.
CC /FtId=VSP_004923.
CC /FtId=VSP_004924.
CC /FtId=VSP_004925.
CC /FtId=VSP_004926.
CC /FtId=VSP_004927.
CC /FtId=VSP_004928.
CC /FtId=VSP_004929.
CC /FtId=VSP_004930.
CC /FtId=VSP_004931.
CC /FtId=VSP_004932.
CC /FtId=VSP_004933.
CC /FtId=VSP_004934.
CC /FtId=VSP_004935.
CC /FtId=VSP_004936.
CC /FtId=VSP_004937.
CC /FtId=VSP_004938.
CC /FtId=VSP_004939.
CC /FtId=VSP_004940.
CC /FtId=VSP_004941.
CC /FtId=VSP_004942.
CC /FtId=VSP_004943.
CC /FtId=VSP_004944.
CC /FtId=VSP_004945.
CC /FtId=VSP_004946.
CC /FtId=VSP_004947.
CC /FtId=VSP_004948.
CC /FtId=VSP_004949.
CC /FtId=VSP_004950.
CC /FtId=VSP_004951.
CC /FtId=VSP_004952.
CC /FtId=VSP_004953.
CC /FtId=VSP_004954.
CC /FtId=VSP_004955.
CC /FtId=VSP_004956.
CC /FtId=VSP_004957.
CC /FtId=VSP_004958.
CC /FtId=VSP_004959.
CC /FtId=VSP_004960.
CC /FtId=VSP_004961.
CC /FtId=VSP_004962.
CC /FtId=VSP_004963.
CC /FtId=VSP_004964.
CC /FtId=VSP_004965.
CC /FtId=VSP_004966.
CC /FtId=VSP_004967.
CC /FtId=VSP_004968.
CC /FtId=VSP_004969.
CC /FtId=VSP_004970.
CC /FtId=VSP_004971.
CC /FtId=VSP_004972.
CC /FtId=VSP_004973.
CC /FtId=VSP_004974.
CC /FtId=VSP_004975.
CC /FtId=VSP_004976.
CC /FtId=VSP_004977.
CC /FtId=VSP_004978.
CC /FtId=VSP_004979.
CC /FtId=VSP_004980.
CC /FtId=VSP_004981.
CC /FtId=VSP_004982.
CC /FtId=VSP_004983.
CC /FtId=VSP_004984.
CC /FtId=VSP_004985.
CC /FtId=VSP_004986.
CC /FtId=VSP_004987.
CC /FtId=VSP_004988.
CC /FtId=VSP_004989.
CC /FtId=VSP_004990.
CC /FtId=VSP_004991.
CC /FtId=VSP_004992.
CC /FtId=VSP_004993.
CC /FtId=VSP_004994.
CC /FtId=VSP_004995.
CC /FtId=VSP_004996.
CC /FtId=VSP_004997.
CC /FtId=VSP_004998.
CC /FtId=VSP_004999.
CC /FtId=VSP_005000.
CC /FtId=VSP_005001.
CC /FtId=VSP_005002.
CC /FtId=VSP_005003.
CC /FtId=VSP_005004.
CC /FtId=VSP_005005.
CC /FtId=VSP_005006.
CC /FtId=VSP_005007.
CC /FtId=VSP_005008.
CC /FtId=VSP_005009.
CC /FtId=VSP_005010.
CC /FtId=VSP_005011.
CC /FtId=VSP_005012.
CC /FtId=VSP_005013.
CC /FtId=VSP_005014.
CC /FtId=VSP_005015.
CC /FtId=VSP_005016.
CC /FtId=VSP_005017.
CC /FtId=VSP_005018.
CC /FtId=VSP_005019.
CC /FtId=VSP_005020.
CC /FtId=VSP_005021.
CC /FtId=VSP_005022.
CC /FtId=VSP_005023.
CC /FtId=VSP_005024.
CC /FtId=VSP_005025.
CC /FtId=VSP_005026.
CC /FtId=VSP_005027.
CC /FtId=VSP_005028.
CC /FtId=VSP_005029.
CC /FtId=VSP_005030.
CC /FtId=VSP_005031.
CC /FtId=VSP_005032.
CC /FtId=VSP_005033.
CC /FtId=VSP_005034.
CC /FtId=VSP_005035.
CC /FtId=VSP_005036.
CC /FtId=VSP_005037.
CC /FtId=VSP_005038.
CC /FtId=VSP_005039.
CC /FtId=VSP_005040.
CC /FtId=VSP_005041.
CC /FtId=VSP_005042.
CC /FtId=VSP_005043.
CC /FtId=VSP_005044.
CC /FtId=VSP_005045.
CC /FtId=VSP_005046.
CC /FtId=VSP_005047.
CC /FtId=VSP_005048.
CC /FtId=VSP_005049.
CC /FtId=VSP_005050.
CC /FtId=VSP_005051.
CC /FtId=VSP_005052.
CC /FtId=VSP_005053.
CC /FtId=VSP_005054.
CC /FtId=VSP_005055.
CC /FtId=VSP_005056.
CC /FtId=VSP_005057.
CC /FtId=VSP_005058.
CC /FtId=VSP_005059.
CC /FtId=VSP_005060.
CC /FtId=VSP_005061.
CC /FtId=VSP_005062.
CC /FtId=VSP_005063.
CC /FtId=VSP_005064.
CC /FtId=VSP_005065.
CC /FtId=VSP_005066.
CC /FtId=VSP_005067.
CC /FtId=VSP_005068.
CC /FtId=VSP_005069.
CC /FtId=VSP_005070.
CC /FtId=VSP_005071.
CC /FtId=VSP_005072.
CC /FtId=VSP_005073.
CC /FtId=VSP_005074.
CC /FtId=VSP_005075.
CC /FtId=VSP_005076.
CC /FtId=VSP_005077.
CC /FtId=VSP_005078.
CC /FtId=VSP_005079.
CC /FtId=VSP_005080.
CC /FtId=VSP_005081.
CC /FtId=VSP_005082.
CC /FtId=VSP_005083.
CC /FtId=VSP_005084.
CC /FtId=VSP_005085.
CC /FtId=VSP_005086.
CC /FtId=VSP_005087.
CC /FtId=VSP_005088.
CC /FtId=VSP_005089.
CC /FtId=VSP_005090.
CC /FtId=VSP_005091.
CC /FtId=VSP_005092.
CC /FtId=VSP_005093.
CC /FtId=VSP_005094.
CC /FtId=VSP_005095.
CC /FtId=VSP_005096.
CC /FtId=VSP_005097.
CC /FtId=VSP_005098.
CC /FtId=VSP_005099.
CC /FtId=VSP_0051
```

FT VARSPLIC 375 738 Missing (in isoform 2).
FT CONFLICT 46 46 /FTID=VSP_004532.
SQ SEQUENCE 738 AA; 82177 MW; 168E8757143BF1F6 CRC64;
Query Match 3.7%; Score 98.5; DB 1; Length 738;
Best Local Similarity 21.4%; Pred. No. 2.5; 99; Indels 69; Gaps 10;
Matches 55; Conservative 34; Mismatches 99;
QY 274 IFFATFCQKN-DIELTPKGYEATFPNENYLEKTSKAAPSRFLNMDCPNHFQKVMKL 332
DB 441 VFKKTENSENLVDLT---YDIQSFTDVTYRQAINSKM-----FELDMKI 482
QY 333 EAVDLMPELRLICVATVRVVRHLISIF-----DGM-----DSEVDQWDCESPD 377
DB 483 AAMHVKKQQL-----HQLLPFVHLQKHKHSTEGVKLALNDSLDLSMDSDNSM 532
QY 378 IYPVGMCELTGYQLQ-----PPVAAEPATPLKAKEAT-----KKKKQFGKRRK 421
DB 533 SVPSPSTAMKTSPLNSSGSGRNSPAPAVTAASVTISQASEVSVPOANSSESFGPFSSE 592
QY 422 RIPTKTRPLRQGGKKLLEDDPGARKISSEVPGEIIAVRVKEHLDV---ASPDKAS 478
DB 593 SIPQTATQPAISPPPKETVSRVSSSTRLVNPSPRPSGNTATKVPNPVGVKRTSSPNKEE 652
QY 479 SPELPVSVENIKQETDD 495
DB 653 SP-----KKTKTEDE 663
RESULT 10
Z198 HUMAN STANDARD; PRT: 1377 AA.
AC Q9UBW7; Q43212; Q43434; Q60898; Q9H538; Q9UEU2;
DT 28-FEB-2003 (Rel. 41; Created)
DT 28-FEB-2003 (Rel. 41; Last sequence update)
DE Zinc finger protein 198 (fused in myeloproliferative disorders
protein) (Rearranged in atypical myeloproliferative disorder protein).
GN ZNF198 OR FIM OR RAMP.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98245146; PubMed=9576949;
RA Popovici C., Adlaide J., Ollendorff V., Chaffanet M., Guasch G.,
RA Jacrot M., Leroux D., Birnbaum D., Pebusque M.-J.;
RT "Fibroblast growth factor receptor 1 is fused to FIM in stem-cell
RT myeloproliferative disorder with t(8;13)(p12;q12).";
RL Proc. Natl. Acad. Sci. U.S.A. 95:5712-5717(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98384238; PubMed=9715603;
RA Reiter A., Schall J., Kulkarni S., Chase A., Macdonald D.H.C.,
RA Aguilar R.C.T., Goncalves C., Hernandez J.M., Jennings B.A.,
RA Goldman J.M., Cross N.C.P.;
RT "Consistent fusion of ZNF198 to the fibroblast growth factor receptor-
RT 1 in the t(8;13)(p11;q12) myeloproliferative syndrome.";
RL Blood 92:1735-1742(1998).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99107818; PubMed=9889006;
RA Kulkarni S., Reiter A.J., Smedley D., Goldman J.M., Cross N.C.P.;
RT "The genomic structure of ZNF198 and location of breakpoints in the
RT t(8;13) myeloproliferative syndrome.";
RL Genomics 55:118-121(1999).
RN [4]
RP SEQUENCE FROM N.A.
RA Blakey S., Wall M.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [5]

RP SEQUENCE OF 152-1377 FROM N.A.
RX MEDLINE=98361795; PubMed=9694738;
RA Still I.H., Cowell J.K.;
RT "The t(8;13) atypical myeloproliferative disorder: further analysis of
RT the ZNF198 gene and lack of evidence for multiple genes disrupted on
RT chromosome 13.";
RL Blood 92:1456-1458(1998).
RN [6]
RP SEQUENCE OF 170-1020 FROM N.A.
RX MEDLINE=98167848; PubMed=9499416;
RA Smedley D., Hamoudi R., Clark J., Warren W., Abdul-Rauf M., Somers G.,
RA Venter D., Fagan K., Cooper C., Shipley J.;
RT "The t(8;13)(p11;q11-12) rearrangement associated with an atypical
RT myeloproliferative disorder fuses the fibroblast growth factor
RT receptor 1 gene to a novel gene RAMP.";
RL Hum. Mol. Genet. 7:637-642(1998).
RN [7]
RP SEQUENCE OF 621-1377 FROM N.A.
RX MEDLINE=98085877; PubMed=9425908;
RA Xiao S., Nalabolu S.R., Aster J.C., Ma J., Abruzzo L., Jaffe E.S.,
RA Stone R., Weissman S.M., Hudson T.J., Fletcher J.A.;
RT "FGFR1 is fused with a novel zinc-finger gene, ZNF198, in the t(8;13)
RT leukaemia/lymphoma syndrome.";
RL Nat. Genet. 18:84-87(1998).
CC -!- FUNCTION: May function as a transcription factor.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- DISEASE: Involved in a t(8;13)(p12;q12) chromosomal translocation
CC which involves FGFR1 AND ZNF198. The resulting transcript is a
CC possible candidate for stem cell leukemia lymphoma syndrome/SCLL.
CC -!- SIMILARITY: Contains 5 MYN-type zinc fingers.
CC -!- CAUTION: Ref.5 sequence differs from that shown due to a
CC frameshift in position 330.
CC -!- CAUTION: Ref.6 sequence differs from that shown due to frameshifts
CC in positions 330, 966, 1009 and 1017.
CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/ZNF198ID114.html".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; Y13472; CAA73875.1; -
CC EMBL; AJ224901; CAA12204.1; -
CC EMBL; AJ007676; CAA07604.1; -
CC EMBL; AJ007677; CAA07604.1; JOINED.
CC EMBL; AJ007678; CAA07604.1; JOINED.
CC EMBL; AJ007679; CAA07604.1; JOINED.
CC EMBL; AJ007680; CAA07604.1; JOINED.
CC EMBL; AJ007681; CAA07604.1; JOINED.
CC EMBL; AJ007682; CAA07604.1; JOINED.
CC EMBL; AJ007683; CAA07604.1; JOINED.
CC EMBL; AJ007684; CAA07604.1; JOINED.
CC EMBL; AJ007685; CAA07604.1; JOINED.
CC EMBL; AJ007686; CAA07604.1; JOINED.
CC EMBL; AJ007687; CAA07604.1; JOINED.
CC EMBL; AJ007688; CAA07604.1; JOINED.
CC EMBL; AJ007689; CAA07604.1; JOINED.
CC EMBL; AJ007690; CAA07604.1; JOINED.
CC EMBL; AJ007691; CAA07604.1; JOINED.
CC EMBL; AJ007692; CAA07604.1; JOINED.
CC EMBL; AJ007693; CAA07604.1; JOINED.
CC EMBL; AJ007694; CAA07604.1; JOINED.
CC EMBL; AJ007695; CAA07604.1; JOINED.
CC EMBL; AJ007696; CAA07604.1; JOINED.
CC EMBL; AL137119; CAC42467.1; -
CC EMBL; AL138688; CAC16956.1; -
CC EMBL; AF060181; AAC23591.1; ALT_FRAME.
CC EMBL; AF035374; AAB88464.1; ALT_FRAME.
CC EMBL; AF012126; AAC01561.1; -

Centromere/microtubule binding protein CBP5 (Centromere-binding factor 5) (Nucleolar protein CBP5).

DE 343 EBDTPRWGLGPIAQQKKQKADKLDKYGR-VNENTPENWKKDYKDL-DEQAPAPPIPE 400

GN CBP5.

OS Candida albicans (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; mitosporic Saccharomycetales; Candida.

OX NCBI_TaxID=5476;

RN [1]

RP SEQUENCE FROM N.A.

RA Jiang W., Clifford J., Koltin Y.;

RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A

CC CENTROMERE DNA-CBP3-BINDING FACTOR AND IS INVOLVED IN MITOTIC

CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN

CC SOME WAY ASSOCIATED WITH THE CBP3 110 kDa SUBUNIT (CBP3A) (BY

CC SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).

CC -!- SIMILARITY: BELONGS TO THE TRUE FAMILY OF PSEUDOURIDINE SYNTHASES.

CC -!- SIMILARITY: Contains 1 PUA domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; U59149; AB894297.1; ..

DR InterPro; IPR004802; Cbf5.

DR InterPro; IPR002478; PUA.

DR InterPro; IPR002501; TruB N.

DR InterPro; IPR004521; Unchar_dom_2.

DR Pfam; PF01472; PUA; 1.

DR Pfam; PF01509; TruB N; 1.

DR SMART; SM00359; PUA; 1.

DR TIGRFAMs; TIGR00425; CBF5; 1.

DR TIGRFAMs; TIGR00451; unchar_dom_2; 1.

DR PROSITE; PS00890; PUA; 1.

KW Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.

FT DOMAIN 267 342

PUA.

SQ SEQUENCE 479 AA; 54321 MW; 3BAF5104E12C9EB6 CRC64;

Query Match 3.4%; Score 92.5; DB 1; Length 479;

Best Local Similarity 20.9%; Pred. No. 4.4;

Matches 103; Conservative 61; Mismatches 155; Indels 173; Gaps 27;

QY 42 NDASHDFWNLGTVDPVHPFGICAINSKILVPPETIHAKFTDWKGYLMKLVGSRTPVDF 101

DB 70 NPSHEV-----VAMI---KRLRVEKTHS-----GTLDPKVTGLVCIIDR 109

QY 102 HIKVSMKYPFRQGMLEVV-----DKSQVSRTRMAYVDVTVIGG----- 141

DB 110 ATRIVKS-----QQGAGKEYVCIVRLHEQLKDDKELNR-----ALENLGALFORPLISA 160

QY 142 -----RLRLIYEDGSDDD-----FWC-----HWSPLIH-----PVGWS-----RVV 174

DB 161 VKQLRVRTVYDSKLIFFDNKRGVLGVFWASCEAGTYMRTLCVHLGMLLVGGVGHMQLRVL 220

QY 175 GHGFKMSERSDMAHPTFRKIVCDAPVLYFKKVRVYVTEGGWFEGMKLEATDPLNLGN 234

DB 221 RSG-AMSESNNVTLHVL-----DA-QYVDYDTR-----DESYLEKTIQPL----- 260

QY 235 ICVATVCKVLLDGLVMTCVDPGSGSTDGLMFCFHASHAIFATFCQKNDIELTPPKGYE 294

DB 261 -----ESLLVGYKRVVW-----KDSAVNSVCYCAK--LIMPGLLRYERGIEL-----YD 302

QY 295 AQTFNWNLYEKTNSKAAPSRLENMDCPNHGFVKGM-KLEAVLMEPRLICVATVKR--- 350

DB 303 -----EVLMTTKGEA-----IAIGQMSTVDLQSCDGHIVAKVRCRM 342

QY 351 -----VVHRLSIHFDGMDSEYDQWDCSDPDYIPVGHCELTGVLOLPP--- 394

Db 343 EBDTPRWGLGPIAQQKKQKADKLDKYGR-VNENTPENWKKDYKDL-DEQAPAPPIPE 400

QY 395 ---VAAPATPLKA-----KEATKKKKQFGKKRIPPTKTRPRRQSGKK 437

DB 401 SKLVAPEQLPKKSLIBEVEDVDVEDSKSKKKKKKKKKKKKKKKKKKKKK 460

QY 438 PLEDDPQDQARK 449

DB 461 RKAEDSSKSEK 472

RESULT 13

AT12 HUMAN

ID AT12 HUMAN STANDARD; PRT; 1593 AA.

AC P58397.

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12).

GN ADAMTS12

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Petal lung;

RX MEDLINE=21264577; PubMed=11279086;

RA Cal S., Argueelles J.M., Fernandez P.L., Lopez-Otin C.;

RT "Identification, characterization, and intracellular processing of ADAM-TS12, a novel human disintegrin with a complex structural organization involving multiple thrombospondin-1 repeats.";

RT J. Biol. Chem. 276:17932-17940(2001).

RL

CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).

CC -!- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely expressed in gastric carcinomas and in cancer cells of diverse origin.

CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).

CC -!- PTM: THE PRECURSOR IS CLEAVED BY A PURIN ENDOPEPTIDASE.

CC -!- PTM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1 DOMAINS.

CC -!- SIMILARITY: Belongs to peptidase family M12B.

CC -!- SIMILARITY: Contains 1 disintegrin-like domain.

CC -!- SIMILARITY: Contains 1 PLAC domain.

CC -!- SIMILARITY: Contains 8 TSP type-1 domains.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AJ250735; CAC20419.1; ..

DR Genew; HGNC:14605; ADAMTS12.

DR MIM; 606184; ..

DR InterPro; IPR001762; Disintegrin.

DR InterPro; IPR002870; Pep M12B propep.

DR InterPro; IPR001590; Reprolysin.

DR InterPro; IPR000884; TSP1

DR InterPro; IPR006025; Zn_Mtpeptdse.

DR Pfam; PF01562; Pep M12B propep; 1.

DR Pfam; PF01421; Reprolysin; 1.

DR Pfam; PF00090; tsp_1; 6.


```

DR SMART; SM00209; TSP1; 8.
DR PROSITE; PS00215; ADAM MEPRO; 1.
DR PROSITE; PS00546; CYSTEINE SWITCH; FALSE NEG.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.
DR PROSITE; PS00214; DISINTEGRIN_2; FALSE NEG.
DR PROSITE; PS00092; TSP1; 6.
DR PROSITE; PS00142; ZINC PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 25 POTENTIAL.
FT PROPEP 26 240 BY SIMILARITY.
FT CHAIN 241 1593 ADAMTS-12.
FT DOMAIN 241 464 METALLOPROTEASE.
FT DOMAIN 465 544 DISINTEGRIN-LIKE.
FT DOMAIN 542 597 TSP TYPE-1 1.
FT DOMAIN 597 700 CYS-RICH.
FT DOMAIN 701 826 SPACER 1.
FT DOMAIN 823 882 TSP TYPE-1 2.
FT DOMAIN 886 942 TSP TYPE-1 3.
FT DOMAIN 943 996 TSP TYPE-1 4.
FT DOMAIN 996 1315 SPACER 2.
FT DOMAIN 1312 1365 TSP TYPE-1 5.
FT DOMAIN 1367 1421 TSP TYPE-1 6.
FT DOMAIN 1422 1470 TSP TYPE-1 7.
FT DOMAIN 1471 1531 TSP TYPE-1 8.
FT DOMAIN 302 305 POLY-GLU.
FT DOMAIN 1538 1570 PLAC.
FT SITE 208 208 CYSTEINE SWITCH (POTENTIAL).
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 393 393 BY SIMILARITY.
FT METAL 396 396 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 105 105 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 125 125 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 215 215 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 685 685 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 790 790 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 951 951 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 1104 1104 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 1275 1275 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 1300 1300 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 1320 1320 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 1371 1371 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 1378 1378 N-LINKED (GLNAC. . .) (POTENTIAL).
FT CARBOHYD 1503 1503 N-LINKED (GLNAC. . .) (POTENTIAL).
SQ SEQUENCE 1593 AA; 177545 MW; 07P9F48E63BD83A3 CRC64;

Query Match 3.4%; Score 92.5; DB 1; Length 1593;
Best Local Similarity 18.4%; Pred. No. 22;
Matches 93; Conservative 59; Mismatches 166; Indels 187; Gaps 26;

QY 9 LNSDAVLPSRYWYIASVIQTAGYRVLLRYSGFENDAS----HDFWCNLG-TVDVHPIGWC 63
DB 378 INDSGLFLAF---TIAHELHSGFGLQHDGKNCDEPVGRHPIYINSRLOQYDPTPLTWS 433
QY 64 AINSKILVPPRTTHAKFTD--WKGYLMKRLVGRSLPVDPHIKOVESMKYPPRQGMRLSV 121
DB 434 KCSEVI-----TRLDKRW-GFCLDDI-----PKKGLKSKV 465
QY 122 V-----DKSQSRTRMAVDTVIGRLRLLYEDGSDDDDFW 157
DB 466 IAPGVYDVHQQLOLYGNPNATFCQEVENVQTLWCSSVKGFCSRKLDAADAGTQCCKW 525
QY 158 C-----HMSPLHPVGMWRVGHGKIMSER-----183
DB 526 CMAGKCIYVKKPESIPGGWGRNSPNH---CSRTGAGVQSAERLNNPPEPKGGKYCT 582
QY 184 -----RSDMAHPTFRKLYC---DAVPY-----LFKKVR---AVYTE- 214
DB 583 GERKRYRLCNVHPCKRSEA---PTFRMQCSEFDTPYKNELYHWFPIFNPAHPCELYCRP 639
QY 215 -GGWFEFGMKLEADPL-----NLGNICVATVCKVLLDGYLMICVDGGPSTDGLDWFY 267

```

```

DB 640 IDGQFSEKMLDAVIDGTPCFEGGNSRNVINCIGICKVWGCDE---IDSNATEDRCGVLG 696
QY 268 HASSHAIFPATFOCKN-----DIELTPPKGYEAQTENWE---NVLEKTKSKAAPSLFNM 319
DB 697 DGSSCQIVRKFKQKSGSYVDGLPKPGARDIRVVEIEGAGNFLAIRSD--PEKY-- 752
QY 320 DCPNHGFKV---GMKLEAVDL-----NEPRLICVATVVRVHRL----- 356
DB 753 --LNGGFIIOWNGNYKLAVTFQYDRKGDLKLMATGPTNESVMIQLLFQVTPNGIKYEY 810
QY 357 SIHEDGWDG-----EYDQWDC 373
DB 811 TIQKGLDNDVEQMYFWQYGHWTCE 835

RESULT 14
H1_MAIZE STANDARD; PRT; 245 AA.
ID_H1_MAIZE
AC P2344;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. W22;
RX MEDLINE=91227140; PubMed=1709276;
RA Rafafmaharitra P.; Chaubet N.; Philipps G.; Gigot C.;
RT "Nucleotide sequence and expression of a maize H1 histone cDNA.";
RL Nucleic Acids Res. 19:1491-1496(1991).
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X57077; CAA40362.1; -.
CC PIR; S26826; S26826.
CC MaizeDB; 25540; -.
CC InterPro; IPR005818; Histone H1/H5.
CC InterPro; IPR005819; Histone_H5.
CC InterPro; IPR003216; Linkerhist_N.
CC Pfam; PF00538; linker_histone_1.
CC PRINTS; PR00624; HISTONEHS.
CC ProDom; PD000373; Linkerhist_N; 1.
CC SMART; SM00526; H15; 1.
CC Chromosomal protein; Nuclear protein; DNA-binding.
FT INIT MET 0 0
FT DOMAIN 1 30 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 31 45 ARG/LYS-RICH (BASIC).
FT DOMAIN 46 121 GLOBULAR.
SQ SEQUENCE 245 AA; 25217 MW; D1FD73DAA74A6778 CRC64;

Query Match 3.4%; Score 91.5; DB 1; Length 245;
Best Local Similarity 33.3%; Pred. No. 2.2;
Matches 34; Conservative 12; Mismatches 37; Indels 19; Gaps 5;

QY 392 QPVAEEP--ATPLKAEATKK-----KKQFGKKRGRIPPTKTRPLRQSK-KPL 439
DB 139 KPAAAKPKAKTPAKKATPKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 198

```

Search completed: February 4, 2004, 09:07:24
Job time : 21 secs

```
QY      440 LEDDPQG-ARKISSEPVPGELIAIVRYKEEHLDVASPKASSP 480
       :   |   |||::|||::|   |||::|   |||::|
Db     199 AKKAGRAAKATSKADTPGK-----KAPAKGAAPSKAATP 234

RESULT 15
EAF3_YEAST
ID    EAF3_YEAST          STANDARD;          PRT;        401 AA.
AC    Q12432;
DT    30-MAY-2000 (Rel. 39, Created)
DD    30-MAY-2000 (Rel. 39, Last sequence update)
DE    28-FEB-2003 (Rel. 41, Last annotation update)
DT     ESA1-associated factor 3.
DN     EAF3 OR YPR023C OR YP9367.03C.
GN     Saccharomyces cerevisiae (Baker's yeast).
OS     Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC     Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI TaxID=4932;
[!]_TaxID=4932;
RN     [!]-SEQUENCE FROM N.A.
RP     SEQUENCE FROM N.A.
RC     STRAIN=S288C / AB972;
RX     MEDLINE=97313271; PubMed=9169875;
RZ     BUSBY H.S., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoerge W.,
RA     Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA     Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA     Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA     Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterloft A.,
RA     Duncan M., Foeth M., Fortin N., Friese J.D., Fritz C., Goffeau A.,
RA     Hall J.J., Hebling U., Heimann K., Halbert H., Hillier L.,
RA     Hunicke-Smith S., Hyman R., Johnston M., Kalman S., Klein K.,
RA     Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA     Marathe R., Messenguy F., Mwewa H.-W., Mirtipati S., Moestl D.,
RA     Mueller-Auer S., Nanath A., Newtich U., Oetner P., Pearson D.,
RA     Petal F.X., Pohl T.M., Punelle D., Schafer M., Schafé M.,
RA     Schrems B., Schramm S., Schroeder M., Sidcu A.M., Tetelin H.,
RA     Urrestazu L.A., Ushinsky S., Vierendeels F., Viessers S., Voss H.,
RA     Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA     Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT     "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL     Nature 387:103-105(1997).
[2]
RP     CHARACTERIZATION.
RX     PubMed=110306083;
RA     Eisen A., Utley R.T., Nourani A., Allard S., Schmidt P., Lane W.S.,
RA     Lucchesi J.C., Cote J.;
RT     "The yeast NuA4 and Drosophila MSL complexes contain homologous
RT     subunits important for transcriptional regulation.";
RJ     J. Biol. Chem. 276:3483-3493(2000).
CC     -!- FUNCTION: MAY PLAY A ROLE IN TRANSCRIPTION REGULATION AS PART OF
CC     THE NUA4 COMPLEX.
CC     -!- SUBUNIT: COMPONENT OF THE NUA4 HISTONE ACETYLTRANSFERASE COMPLEX.
CC     INTERACTS WITH ESA1.
CC     -!- SURCELLULAR LOCATION: Nuclear (Probable).
CC     -!- SIMILARITY: BELONGS TO THE WRG FAMILY.
CC     -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC     EMBL; 271255; CAA95019.1; -.
DR     EMBL; Z49274; CAA89277.1; -.
DR     PIR; S54497; S54497.
DR     SGD; S0006227; EAF3.
DR     GO; GO:0000123; C:histone acetyltransferase complex; IDA.
DR     GO; GO:0004402; P:histone acetyltransferase activity; IDA.
DR     GO; GO:0016573; P:histone acetylation; IDA.
DR     GO; GO:0006357; P:regulation of transcription from Pol II pro..; IMP.
DR     InterPro; IPR000953; Chromo.
```

2PR 14-SEP-2000; 2000US-0662191.
2PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Aundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 FI Zhao QA, Zhou F, Goodrich R, Drmanac RT;
 XX WPI: 2001-442253/47.
 DR N-PSDB; AAI58732.
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX Example 4; SEQ ID NO 2721; 10078pp; English.
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAI38642-AAI42213) with neurotrophic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX SQ Sequence 495 AA;
 Query Match 100.0%; Score 2683; DB 22; Length 495;
 Best Local Similarity 100.0%; Pred. No. 1.5e-269;
 Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MKGMKVEVLSNDVLPBRYVWIASVIOAGYRVLLRYEGFENDASHDFWNCNLTGTVDVHPI 60
 Db 1 MKGMKVEVLSNDVLPBRYVWIASVIOAGYRVLLRYEGFENDASHDFWNCNLTGTVDVHPI 60
 Qy 61 GWCAINSKILVPPRTIHAKTDMKGYLMKRLVGSRTLPVDFHFKMVESKKYPPRQGMRL 120
 Db 61 GWCAINSKILVPPRTIHAKTDMKGYLMKRLVGSRTLPVDFHFKMVESKKYPPRQGMRL 120
 Qy 121 VVDKSOVSRTMAVVDTVIGRLRLLYEDGSDDDFWCHWMSPLIHPVGMSSRVVGHGIM 180
 Db 121 VVDKSOVSRTMAVVDTVIGRLRLLYEDGSDDDFWCHWMSPLIHPVGMSSRVVGHGIM 180
 Qy 181 SERSDMAHPTFRKIYCDVAVPLFKKRVAVYTEGGFEGMKLEAIDPLNLGNICVATV 240
 Db 181 SERSDMAHPTFRKIYCDVAVPLFKKRVAVYTEGGFEGMKLEAIDPLNLGNICVATV 240
 Qy 241 CKVLLDGYLMICVDGSGSTGLDWFVCHASSHAIFPATFCQKNDIELTPPKGYEATFNW 300
 Db 241 CKVLLDGYLMICVDGSGSTGLDWFVCHASSHAIFPATFCQKNDIELTPPKGYEATFNW 300
 Qy 301 ENYLEKTSKAAPRLNMDCPNPHGFKVGMKLEAVDLMEPRLLICVATKRVVHRLLSIH 360
 Db 301 ENYLEKTSKAAPRLNMDCPNPHGFKVGMKLEAVDLMEPRLLICVATKRVVHRLLSIH 360
 Qy 361 DGMSEVDQWDCSPPIYPVWCELTGYQLQPPVAEPATPLKAEATKKKKQFGKKR 420
 Db 361 DGMSEVDQWDCSPPIYPVWCELTGYQLQPPVAEPATPLKAEATKKKKQFGKKR 420
 Qy 421 KRIPTKTRPLRQSKKPLLEDDPQGARKISSEVPGEIIVAVKGEHLDPVSPDKASSP 480
 Db 421 KRIPTKTRPLRQSKKPLLEDDPQGARKISSEVPGEIIVAVKGEHLDPVSPDKASSP 480
 Qy 481 ELPVSVENIKQETDD 495
 |||||

Db 481 ELPVSVENIKQETDD 495
 RESULT 2
 AAB60488
 ID AAB60488 standard; Protein; 495 AA.
 XX
 AC AAB60488;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Human cell cycle and proliferation protein CCYPR-36, SEQ ID NO:36.
 KW Cell cycle and proliferation protein; CCYPR; human; agonist;
 KW antagonist; gene therapy; detection; gene therapy;
 KW transgenic animal disease model; immune disorder;
 KW developmental disorder; cell signalling disorder;
 KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;
 KW arteriosclerosis; asthma; allergy; diabetes mellitus;
 XX menstrual cycle disorder; bacterial infection.
 OS Homo sapiens.
 XX WO200107471-A2.
 XX 01-FEB-2001.
 XX 21-JUL-2000; 2000WO-US19948.
 PR 21-JUL-1999; 99US-0145075.
 PR 08-SEP-1999; 99US-0153129.
 PR 10-NOV-1999; 99US-0184647.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;
 PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;
 XX WPI: 2001-112727/12.
 DR N-PSDB; AAF59625.
 XX Human cell cycle and proliferation proteins and polynucleotides are
 PT used to treat, diagnose and prevent immune, developmental and cell
 PT signalling disorders and cell proliferative disorders including cancer -
 XX Claim 1; Page 142-143; 205pp; English.
 CC Sequences AAB60453-AAB60506 represent 54 human cell cycle and
 CC proliferation proteins (CCYPR), which are encoded by AAF59590-AAF59643.
 CC CCYPR and agonists of CCYPR are used to treat diseases or conditions
 CC associated with decreased expression of functional CCYPR, while CCYPR
 CC antagonists are used to treat diseases or conditions associated with
 CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
 CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
 CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect
 CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)
 CC that specifically bind to CCYPR, and in drug screening methods to
 CC identify compounds that modulate the activity of CCYPR. CCYPR
 CC nucleotides can be used to generate transgenic animal models of human
 CC disease, and can be used in gene therapy in target cells with genetic
 CC abnormalities with respect to the expression of CCYPR for the
 CC treatment or prevention of a disorder associated with CCYPR.
 CC Diseases which can be diagnosed, treated and prevented using CCYPR
 CC proteins, nucleic acids, agonists or antagonists include immune,
 CC developmental and cell signalling disorders, and cell proliferative
 CC disorders including cancer. Specific examples of these disorders
 CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
 CC diabetes mellitus, disorders of the menstrual cycle and infections
 CC caused by bacteria.
 XX Sequence 495 AA;
 Query Match 100.0%; Score 2683; DB 22; Length 495;

ID ABB05680 standard; Protein; 617 AA.
 AC ABB05680;
 DT 30-APR-2002 (first entry)
 DE Human transmembrane protein clone amy2_1114.
 DE Human; foetal brain; foetal kidney; melanoma; testis; amygdala;
 DE gene therapy; chromosome 22q13.31-13.33.
 DE Homo sapiens.
 OS WO200198454-A2.
 PN 27-DEC-2001.
 PD XX
 XX 25-APR-2001; 2001WO-1B02050.
 XX 25-APR-2000; 2000US-199380P.
 XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
 PA Wiemann S;
 PI WPI; 2002-0558660/07.
 DR N-PSDB; ABA93718.
 DR Human cDNA sequences and clones derived from human fetal brain, fetal
 PT kidney, melanoma, testis and amygdala cDNA libraries, useful in genetic
 PT screening and therapy -
 XX Claim 1; Page 130; 611pp; English.
 PS The present invention describes assemblies and computer readable media
 CC comprising novel human cDNA sequences and clones derived from human
 CC foetal brain, foetal kidney, melanoma, testis and amygdala cDNA
 CC libraries. ABA93702 to ABA93766 represent human cDNA sequences from the
 CC present invention which encode the proteins given in ABB05662 to
 CC ABB05729. The human cDNA sequences and clones can be used in gene
 CC therapy. The clones may be used in a variety of applications, for
 CC example they may be used in profiling assays, for providing large arrays
 CC of human genetic material for implementing large-scale screening
 CC strategies and for treating diseases via gene therapy procedures.
 XX Sequence 617 AA;
 SQ
 Query Match 81.4%; Score 2185; DB 23; Length 617;
 Best Local Similarity 100.0%; Pred. No. 1.2e-217;
 Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKGMKVEVLNSDAVLPSRVYWTASVITQAGYRVLLRYEGFENDASHDFWNLGTVDVHPI 60
 DB 211 MKGMKVEVLNSDAVLPSRVYWTASVITQAGYRVLLRYEGFENDASHDFWNLGTVDVHPI 270
 QY 61 GWCAINSKILVPRTHAKFTDWMKYLKRLVGSRTLPVDFHFKWVESMKYFFRQGMRL 120
 DB 271 GWCAINSKILVPRTHAKFTDWMKYLKRLVGSRTLPVDFHFKWVESMKYFFRQGMRL 330
 QY 121 VDKSQVSRTRMAVDTVIGGRLLLYEDGSDDDFWCHWNSPLIHPVGSRRVGHGIM 180
 DB 331 VDKSQVSRTRMAVDTVIGGRLLLYEDGSDDDFWCHWNSPLIHPVGSRRVGHGIM 390
 QY 181 SERSDMAHPTFRKLYCDAPVPLFKKRAVYTEGGWFEGMKLEAIDPLNLGNTCVATV 240
 DB 391 SERSDMAHPTFRKLYCDAPVPLFKKRAVYTEGGWFEGMKLEAIDPLNLGNTCVATV 450
 QY 241 CKVLLDGLYMICVGGPSTDLGDFWFCYHASSHAIFPATFCQKNDIELTPPKYEATFNW 300
 DB 451 CKVLLDGLYMICVGGPSTDLGDFWFCYHASSHAIFPATFCQKNDIELTPPKYEATFNW 510
 QY 301 ENYLEKTSKAAPSRLFNMDCPNHGFKVGMKLEAVDLMEPRLICVATKRVVHRLLSIHF 360

Db 511 ENYLEKTSKAAPSRLFNMDCPNHGFKVGMKLEAVDLMEPRLICVATKRVVHRLLSIHF 570
 QY 361 DGMWSEYDQWDCESPDYIPVGVWCELTYGLOPPVAA 397
 Db 571 DGMWSEYDQWDCESPDYIPVGVWCELTYGLOPPVAA 507
 RESULT 5
 ABB89448
 ID ABB89448 standard; Protein; 329 AA.
 XX AC ABB89448;
 XX 24-MAY-2002 (first entry)
 XX Human polypeptide SEQ ID NO 1824.
 XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 XX anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; anti-ulcer;
 XX vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 XX neurological disease; infection; human; secreted protein.
 OS Homo sapiens.
 XX WO200190304-A2.
 XX 29-NOV-2001.
 XX 18-MAY-2001; 2001WO-US16450.
 XX 19-MAY-2000; 2000US-205515P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Birse CE, Rosen CA;
 XX WPI; 2002-122018/16.
 XX N-PSDB; ABL89857.
 DR Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders -
 XX Claim 11; SEQ ID NO 1824; 2081pp + Sequence Listing; English.
 PS The invention relates to novel genes (ABL89449-ABL90853) and proteins
 CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 329 AA;
 SQ
 Query Match 62.9%; Score 1688; DB 23; Length 329;
 Best Local Similarity 99.0%; Pred. No. 2.1e-166;
 Matches 308; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MKGMKVEVLNSDAVLPSRVYWTASVITQAGYRVLLRYEGFENDASHDFWNLGTVDVHPI 60

Db 1 MKGMKVEVLNSDAVLSRVVWIASVIQTAGYRVLRYEGFENDASHDFWNLGTVDVHPH 60
Qy 61 GWCAINSKILVPPRTTHAKFTDHWKYLKMLVGSRTLVDVFIKMWESMKYPRQGMRL 120
Db 61 GWCAINSKILVPPRTTHAKFTDHWKYLKMLVGSRTLVDVFIKMWESMKYPRQGMRL 120
Qy 121 VDKSQVSRTRMAVDTVIGGRLLRYEDGSDDDDFWCHMWSPLIHPVGSRRVGHGIM 180
Db 121 VDKSQVSRTRMAVDTVIGGRLLRYEDGSDDDDFWCHMWSPLIHPVGSRRVGHGIM 180
Qy 181 SERSDMAHHTPRKIYCDAPVFLFKKRAVYITGGWFEFGMKLEAIDPLNLGNTCVATV 240
Db 181 SERSDMAHHTPRKIYCDAPVFLFKKRAVYITGGWFEFGMKLEAIDPLNLGNTCVATV 240
Qy 241 CKVLLDGYLMICVDGSPSTDLDFWCHWSPLIHPVGSRRVGHGIM 300
Db 241 CKVLLDGYLMICVDGSPSTDLDFWCHWSPLIHPVGSRRVGHGIM 300
Qy 301 ENYLEKTSKA 311
Db 301 ENYLEKTSKA 311

RESULT 6

AA79183
ID AAY79183 standard; Protein; 631 AA.
AC AAY79183;
DT 19-JUN-2000 (first entry)
XX Haematopoietic stem cell specific protein.
DE Haematopoietic stem cell; immune system disorder;
KW leukaemia; antileukaemic; immunomodulator; therapy; mouse.
XX Mus musculus.
XX Key Location/Qualifiers
FT Misc-difference 497 /note= "encoded by TTA"
FT Misc-difference 523 /note= "encoded by CAT"
FT Misc-difference 548 /note= "encoded by TAT"
XX WO200011168-A2.
XX 02-MAR-2000.
XX 20-AUG-1999; 99WO-US19052.
XX 21-AUG-1998; 98US-0138132.
XX (UYPR-) UNIV PRINCETON.
XX Lemischka I, Moore K;
XX WPI; 2000-237650/20.
XX N-PSDB; AAZ94124.
XX Haematopoietic stem cell signaling proteins modulating replication and
XX differentiation for treating immune system disorders and leukaemia -
XX Claim 21; Page 229-231; 256pp; English.

XX The present sequence is that of a mouse haematopoietic stem cell
XX (HSC) specific protein. It is an example of claimed HSC-specific
XX proteins (see AAY79176-93) predicted from novel isolated HSC-specific
XX nucleic acids (see AAY294077-131). The HSCs are especially primitive
XX HSCs (PHSCs) such as unilobal cord cells, bone marrow cells and
XX foetal liver cells. The encoded proteins are growth factors,

CC transcription factors, splicing factors, capping factors, transport
CC proteins, translation factors or replication factors that modulate
CC HSC activity, especially differentiation or replication. The
CC invention provides claimed methods: for identifying PHSC-specific
CC nucleic acids; for generating a stem cell/progenitor cell from
CC PHSCs; for identifying the presence of a PHSC in a sample; for
CC identifying the presence in a sample of a compound that modulates
CC HSC activity; for using such a compound to treat an immune system
CC condition, especially leukaemia; for introducing exogenous nucleic
CC acid into a HSC; and for ex vivo expansion of HSCs. Also claimed
CC are vectors, host cells, and an antibody that specifically binds a
CC an HSC-specific protein.
XX
SQ Sequence 631 AA;

Query Match 52.6%; Score 1410; DB 21; Length 631;
Best Local Similarity 56.2%; Pred. No. 4.6e-137;
Matches 263; Conservative 67; Mismatches 122; Indels 16; Gaps 4;
Qy 2 KGMKVEVLNSDAVLSRVVWIASVIQTAGYRVLRYEGFENDASHDFWNLGTVDVHPH 61
Db 177 ENRIEVPNTDCSLPTKVFVIAGIILKAGYNALLRYEGFENDSLDFWNLGTVDVHPH 236
Qy 62 WCAINSKILVPPRTTHAKFTDHWKYLKMLVGSRTLVDVFIKMWESMKYPRQGMRL 121
Db 237 WCAINSKILVPPRTTHAKFTDHWKYLKMLVGSRTLVDVFIKMWESMKYPRQGMRL 121
Qy 122 VDKSQVSRTRMAVDTVIGGRLLRYEDGSDDDDFWCHMWSPLIHPVGSRRVGHGIM 180
Db 297 VDKRHLCTRVAVVSIGGRLLRYEDGSDDDDFWCHMWSPLIHPVGSRRVGHGIM 356
Qy 181 SERSDMAHHTPRKIYCDAPVFLFKKRAVYITGGWFEFGMKLEAIDPLNLGNTCVATV 240
Db 357 SDITKQDGH-----FDPPLFLFAKVEVDGSEFGMKLEAIDPLNLGNTCVATV 403
Qy 241 CKVLLDGYLMICVDGSPSTDLDFWCHWSPLIHPVGSRRVGHGIM 300
Db 410 RKVLADGFLMIGDSEADGSDWFCVHATSPSIFPVFGCEINNIETLTPRGYTKLPFKW 469
Qy 301 ENYLEKTSKAAPSRLEFNMDCPNHGFVKYKLEAIDPLNLGNTCVATV 360
Db 470 FDYLRGTGSAAPVFLFKKRAVYITGGWFEFGMKLEAIDPLNLGNTCVATV 529
Qy 361 DGWSEYDQWVDCSPDIYVPGWCELTYQLQPPVAAEPATPLKAKEATKK-KKQFGKK 419
Db 530 DGWSEYDQWVDCSPDIYVPGWCELTYQLQPPVAAEPATPLKAKEATKK-KKQFGKK 589
Qy 420 RKRIPPTKTRPQSGKKPLEDDPQARKISSEPVGEIIVRVKKE 467
Db 590 HKKMTTSQ-----KEELLGDYSLFHGASDQESNGSATVYIKOE 630

RESULT 7

ABU03557
ID ABU03557 standard; Protein; 410 AA.
XX ABU03557;
XX AC ABU03557;
XX 21-JAN-2003 (first entry)
XX DE Angiogenesis-associated human protein sequence #102.
XX KW Human; angiogenesis-associated transcript; angiogenesis;
XX KW angiogenesis-associated disease; cancer; cytostatic.
XX OS Homo sapiens.
XX PN WO200279492-A2.
XX PD 10-OCT-2002.
XX PF 14-FEB-2002; 2002WO-US04915.
XX

PR 14-FEB-2001; 2001US-0784356.
 PR 22-FEB-2001; 2001US-0791390.
 PR 19-APR-2001; 2001US-285475P.
 PR 03-AUG-2001; 2001US-310025P.
 PR 13-NOV-2001; 2001US-350666P.
 PR 29-NOV-2001; 2001US-334244P.
 XX (EOSB-) EOS BIOTECHNOLOGY INC.
 XX PA
 XX Murray R, Glynn R, Watson SR, Aziz N;
 XX WPI; 2003-040681/03.
 DR N-PSDB; ABX08841.
 XX
 XX Detecting angiogenesis-associated transcript in a cell for diagnosing
 PT and treating cancer by contacting a sample with a polynucleotide that
 PT exhibits changes in expression level as a function of time in tissue
 PT undergoing angiogenesis -
 XX
 XX Example 2; Page 275; 291pp; English.
 XX
 XX The present invention relates to methods and compositions for
 CC detecting an angiogenesis-associated transcript in a cell in
 CC a patient. The method involves contacting a biological sample from
 CC the patient with a polynucleotide that selectively hybridizes to a
 CC sequence at least 80% identical to any of the angiogenesis-associated
 CC human polynucleotide sequences given in the specification. These
 CC angiogenesis-associated polynucleotide sequences comprise genes that
 CC exhibit changes in expression levels as a function of time in tissue
 CC of the invention are useful for diagnosing and treating angiogenesis
 CC and angiogenesis-associated diseases e.g. cancer. The polynucleotide
 CC sequences are also useful in the gene therapy of such disorders. The
 CC angiogenesis-associated proteins encoded by the polynucleotide
 CC sequences are useful as a vaccine for therapeutic and prophylactic
 CC immunisation. ABU03456-ABU03569 represent angiogenesis-associated
 CC protein sequences.
 XX
 XX Sequence 410 AA;
 SQ
 Query Match 52.0%; Score 1395; DB 24; Length 410;
 Best Local Similarity 62.9%; Pred. No. 8.5e-136;
 Matches 249; Conservative 55; Mismatches 84; Indels 8; Gaps 2;
 QY 2 KGMKVEVLSDAVLPSRVYVIASTVQAGYRVLLRYEGFENDASHDFWNLGTVDVHPIG 61
 Db 10 ENVRVEVPNTDCLPTKFWIAGIYKLVKLTGAKTLPDFSKVSESMQYFFKCMRVEV 69
 QY 62 WCAINSKILVPPRTTHAKFTDWKGYLMKRLVGSRTLPVDFFHKVSESMQYFFKCMRLEV 121
 Db 70 WCAASGKPLVPPRTTHAKFTDWKGYLMKRLVGSRTLPVDFFHKVSESMQYFFKCMRVEV 129
 QY 122 VDKSQSVTRMAVDTVIGRRLRYEDG-DDDDFWCHMSPILHPVGNRRVGHGKIM 180
 Db 130 VDKRLHLCRTVAVDSVIGRRLRYEEDRTDDFWCHMSPILHGHGSRSGHGFKR 189
 QY 181 SRRSDMAHPTFRKTYCDAPYLPFKVRAVYTEGGWFEKGKLEAIDPLNLGNICVATV 240
 Db 190 SDITKKQDGH-----FDTPHLPFAKVEVDQSENFEGKLEAIDPLNLGICVATI 242
 QY 241 CKVLDDGYLMICVDGSPSTDGLDFWCHYHASSHAIPATFCQNDIELTPPKGYEAQTNW 300
 Db 243 RKVLADGFLMIGIDGSEADGDFWCHYHASSHAIPATFCQNDIELTPPKGYEAQTNW 302
 QY 301 ENYLEKTSKAAPSRLENNMDCNFGKVKGMKLEAVDLMEPLRICVATVKKRVVHRLLSHF 360
 Db 303 FYLRGTSIAAPVKLVKNDVNNHGFVGMKLEAVDLMEPLRICVATVTRIIHRLLRHF 362
 QY 361 DQWDSYDQWDCESPDYIPVGCVELTGYQLQPPVA 396
 Db 363 DQWSEYDQWDCESPDYIPVGCVELTGYQLQPPAS 398

RESULT 8
 ABB78695
 ID ABB78695 standard; Protein; 410 AA.
 XX
 AC ABB78695;
 XX
 DT 17-JUL-2002 (first entry)
 XX
 DE Human hsp BAA90919 protein sequence SEQ ID NO:68.
 XX
 KW Caenorhabditis elegans; lin-8; lin-56; lin-61; tumour suppressor;
 KW cell proliferation; nematode; cancer; mutant.
 XX
 OS Homo sapiens.
 XX
 PN WO2001194545-A2.
 XX
 DD 13-DEC-2001.
 XX
 PF 01-JUN-2001; 2001WO-US17909.
 XX
 PF 02-JUN-2000; 2000US-208802P.
 XX
 PA (WASI) MASSACHUSETTS INST TECHNOLOGY.
 XX
 XX Horvitz HR, Davison EM, Lu X;
 XX WPI; 2002-401590/43.
 XX
 PT New Caenorhabditis elegans nucleic acid involved in tumor suppressor
 PT pathway, encoding LIN-8, LIN-56 or LIN-61 polypeptide, useful for
 PT diagnosing and treating cell proliferative diseases such as cancer -
 XX
 PS Disclosure; Fig 8; 116pp; English.
 XX
 XX The present invention describes a substantially pure nematode C. elegans
 CC (Caenorhabditis elegans) nucleic acid (I) encoding a LIN-8 polypeptide
 CC (see ABB78648), LIN-56 polypeptide (see ABB78649), or LIN-61 polypeptide
 CC (see ABB78650), where the polypeptides comprise at least 130, 110, 130
 CC contiguous amino acids of the 386, 322 and 498 amino acid sequences given
 CC in ABB78648 to ABB78650, and modulate cell proliferation. (I) has
 CC cytostatic activity, and can be used in gene therapy. (I) is useful for
 CC modulating proliferation of a cell, and for identifying the compound that
 CC modulates cell proliferation. (I) can be used for diagnosing an animal
 CC (preferably, human) for the presence of the cell proliferation disease,
 CC or an increased chance of developing the disease, by measuring lin-8,
 CC lin-56 or lin-61 nucleic acid expression in a sample obtained from the
 CC animal, by measuring LIN-8, LIN-56 or LIN-61 polypeptide in the sample
 CC using Ab or measuring amount of lin-8, lin-56, lin-61 mRNA in the sample.
 CC (I) is also useful for identifying the compound that modulates cell
 CC proliferation. (I) and its mutants are useful for understanding cell
 CC proliferative diseases including cancer, as well as in diagnosing and
 CC treating cell proliferative diseases. (I) is also useful for identifying
 CC tumour suppressors in other species such as mammals and may be used to
 CC identify therapeutic compounds. The present sequence represents a human
 CC hsp (heat shock protein) protein which is given in comparison with the
 CC C. elegans LIN-61 protein in the exemplification of the present
 CC invention.
 XX
 SQ Sequence 410 AA;
 Query Match 51.7%; Score 1387; DB 23; Length 410;
 Best Local Similarity 62.6%; Pred. No. 5.7e-135;
 Matches 248; Conservative 56; Mismatches 84; Indels 8; Gaps 2;
 QY 2 KGMKVEVLSDAVLPSRVYVIASTVQAGYRVLLRYEGFENDASHDFWNLGTVDVHPIG 61
 Db 10 ENVRVEVPNTDCLPTKFWIAGIYKLVKLTGAKTLPDFSKVSESMQYFFKCMRVEV 69
 QY 62 WCAINSKILVPPRTTHAKFTDWKGYLMKRLVGSRTLPVDFFHKVSESMQYFFKCMRLEV 121
 Db 70 WCAASGKPLVPPRTTHAKFTDWKGYLMKRLVGSRTLPVDFFHKVSESMQYFFKCMRVEV 129

QY 122 VDKQVQRTMAVVDVTIGRRLRLLYEDG-DSDDDFWCHMWSPLIHPVGSRRVGHGIM 180
 DB 130 VDKHLQRTVAVVESVIGRRLRLVYBESDRTDDFWCHMESPLIHHIGWSRSGHFRFR 189
 QY 181 SERSDMAHHTFRKIYCDVAVPLPKVAVYTEGGHFEFGMKLEADPLNLGNICVATV 240
 DB 190 SDITKQDGHFT-----DPPHLFAKVKEVDQSGGEWFKGEMKLEADPLNLSTICVATI 242
 QY 241 CKVLLDGLYLMICVDGSPSTDLGDFWCHMWSPLIHPVGSRRVGHGIM 300
 DB 243 KRVLADGLMIGDSEADGSDWFCVHATSPSIFPVGCEINMIELTPRGYTKLPFKW 302
 QY 301 ENYLEKTSKAAPRLNMDPCNHPGFKVGMKLEADPLNLGNICVATV 360
 DB 303 PDYURETGSTAAPVYKLEKNDVFNHGFVGMKLEADPLNLGNICVATV 362
 QY 361 DGMDSYDQWDCESPDIYPVGMCELTGYQLQPPVA 396
 DB 363 DGWEEYDQWDCESPDIYPVGMCELTGYQLQPPAS 398

RESULT 9

ID ABB70683 standard; Protein; 891 AA.
 AC ABB70683;
 XX
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 38841.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX Drosophila melanogaster.
 OS
 FN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 PA (PEKE) PE CORP NY.
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL14786.

PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
 XX
 PS Disclosure; SEQ ID NO 38841; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 891 AA;

Query March 41.4%; Score 1111; DB 22; Length 891;

Best Local Similarity 43.4%; Pred. No. 9.4e-106;
 Matches 234; Conservative 75; Mismatches 162; Indels 68; Gaps 14;
 QY 3 GMKVEVLNSD-----AVLPSRV---YMTASVQTAGVEVLLRYEGFENDASHDPCNLTGT 54
 DB 218 GMKVEVENTDCDSIEVIQFGQTPTSFVATILEIKYKALMSYEGFDTD-SHDFWNLCN 276
 QY 55 VDVHPIGHCAINSKILVPPRTIHAFTDKWGLYMKRLVGSRTLPVDHFHKVMSMKYFPR 114
 DB 277 AEVHVGWGCATRGKPLIPRTIEHKYKDWKDFLVGRLSGARTLPSNFYKINDLSQSRFR 336
 QY 115 QGMRLEVVDKQSVSRTEMAVVDVTIGRRLRLLYEDGSDDDDFWCHMWSPLIHPVGSRRV 174
 DB 337 LGLNLECDVKORISQVRLATVTIKVKKLFLRY--FDSDDGFWCHEDSPIIHPVGWATTV 394
 QY 175 GHGI-----KMSERRSDMAHHTFRKIYCDVAVPLFK---KVRVAVYTEG--GWPEE 220
 DB 395 GHNLAAPODYLERMLAGREAMEVHEH-----DATIELFKMNTTFDEYSDGKTNFVE 448
 QY 221 GMKLEAIDPLNLGNICVATVCKVLLDGLYLMICVDG-GEPTDGLDWFCHMWSPLIHPV 279
 DB 449 GMKLEAVDPLNLSSICPATVMAVLKFGYMMIRIDSYPQDASGSDWFCYHEKSPCIFPAGF 508
 QY 280 CQKNDIELTPPKGYEAOFTFNWENYLEKTSKAAPRLNMDPCNHPGFKVGMKLEADPLME 339
 DB 509 CSVNNISVTTPNGYDSRTFTWEGYLRDTGAVAAGQHLFHRIPDHGFEVGSLECADLMD 568
 QY 340 PLICVATVTKRVVHRLLSIHFDGWDSEYDQWDCESPDIYPVGMCELTGYQLQPP----- 394
 DB 569 PRLVCVATVARVVGRLKLVHFDGWDTEYDQWDCESADIIYPVGMCELVNHHKLEGPVRAH 628
 QY 395 -VAPAPATPLKAKEATKKKQFGKRRIPPTKTRPLRQSGKPELLEDDDPQGARKISSE 453
 DB 629 QQAPKAPKPKIQRKPKKGAAGK-----TPTDNNTQSVKSRITALKTTTH----- 676
 QY 454 PVPGEIIAIVRVKEEHLDA-----SPDKASSPELF-----VSVENIKQETDD 495
 DB 677 -LPKLSIKLELKPHEHNAAFYENNOPEEGDEEDPDADGDGSGTSHISEQSTTQSSSD 734

RESULT 10

ABB66759
 ID ABB66759 standard; Protein; 1243 AA.
 XX
 AC ABB66759;
 XX
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 27069.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX Drosophila melanogaster.
 OS
 FN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 PA (PEKE) PE CORP NY.
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR WPI; 2001-656860/75.
 DR N-PSDB; ABL10862.
 PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell

QY 465 KEEHLDVAVP--DKASSPELPVSV 486
DB 506 YFNERCFSGPYLNKGRIAEFPQCV 529

RESULT 12

ABP69567
ID ABP69567 standard; Protein; 894 AA.

AC ABP69567;
XX

DT 20-JAN-2003 (first entry)
XX

DE Human polypeptide SEQ ID NO 1614.
XX

KW Human; genome mapping; gene therapy; food supplement; virus; fungus;
cell-proliferative disorder; neurodegenerative disease; bacterial;
Parkinson's disease; Alzheimer's disease; autoimmune disease;
multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
haemostatic; vulnary; fungicide; antibacterial; virucide; protozoacide;
antiarthritic.

OS Homo sapiens.
XX

PN WO200270539-A2.
XX

PD 12-SEP-2002.
XX

PF 05-MAR-2002; 2002WO-US05095.
XX

PR 05-MAR-2001; 2001US-0799451.
XX

FA (HYSE-) HYSEQ INC.
XX

PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI Wehrman T, Wang J, Wang D, Drmanac RT;
XX

DR WPI; 2002-759812/82.
XX

DR N-PSDB; ABZ11784.
XX

PT New polynucleotides comprising sequences assembled from expressed
sequence tags (ESTs), useful for treating cell-proliferative,
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT platelet or coagulation disorders -
XX

PS Claim 9; SEQ ID NO 1614; 1012pp + Sequence Listing; English.
XX

XX The invention relates to an isolated polynucleotide (I) comprising a
nucleotide sequence selected from any of 948 sequences
(ABZ11119-ABZ12066) or their mature protein coding portion, active domain
coding protein or complementary sequences. The polynucleotides are useful
for identifying expressed genes or for physical mapping of human genome.
The encoded polypeptides (ABP68902-ABP69849) are useful as molecular
weight markers, as a food supplement, for generating antibodies, in
medical imaging, screening and diagnostic assays and for treating
cell-proliferative disorders (cancer), neurodegenerative diseases
(Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
sclerosis, diabetes, lupus), genetic disorders, myeloid or lymphoid
disorders, platelet or coagulation disorders, wound, burns, incision,
ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
parasitic), arthritis, etc.

CC Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 894 AA;
XX

Query Match 22.8%; Score 611; DB 23; Length 894;
Best Local Similarity 34.18; Pred. No. 8.8e-54;
Matches 156; Conservative 65; Mismatches 170; Indels 66; Gaps 15;

QY 3 GNMKEVLNSDAVLPSSRVYMIASVIQTAGYRVLLRYEGFENDASHDFWNCNLGTVDVHPIGW 62
DB 77 GNMKEVLANKN--NPDTYVWATIIITTCGQLLLRLCYGEGEDRADFDWCDVVIADLHPVGW 133
QY 63 CAINSKILVPPRTIHAKEKTDWKGILMKELVGSRTLPVDFHIMKVESMKYKPPR----- 114
DB 134 CTQNNKVLMPDPAIKEKYTDWTEFLIRDLTSGRTAPANL-----LEGPLRGKGPIDLI 186
QY 115 -QGMRELVVDKSVSRTRMAVVDTVIGRLRLLY---EDGSDDDDFWCHMSPSLIHPVGW 170
DB 187 TVGSLIELODSQNPFPQYWIIVSVIENVGRLRLRYVGLDTSYDQ-WLFYLDYRLRLPVGM 245
QY 171 SRVGHGIMKERRSD-----MAHH--PTRKIYCDVAPY-----LFKVRVAVTE 214
DB 246 COE-----NKYRMDPPSEIYPLKMASEWKTLEKSLIDAAPFELPMEVFKDHADLRS 298
QY 215 GGMFEKGKLEAIDPLNLGNTCVATCKVLLDGYLMICVDG-GPSTDGLDFWCVHASSHA 273
DB 299 --FFTVGMKLETVANMCEPFYISPASVTKVFNHFFQVITIDLRPEPSKLSMLC-HADSLG 355
QY 274 IFFTFCQNDIELTPPKGYBAQTFNWNYLEKTSKAAAPSLRNMDCPNHGFKVGMKLE 333
DB 356 ILPVQWCLKNGVSLTPPKGYSGQDFWADYHKHGAQEAAPPFCFRNTSFSRGFTKMKLE 415
QY 334 AVDLMEPRLICVATKYRVVHRLLSLTHFDGWSEYDQ-WVDCESPDYYPVGMCELTGYQLQ 392
DB 416 AVNPRNGELCVASVSVKGRMLMLHLEGLQTFVPEVIVDVESMDIFFVGMCEANSY--- 472
QY 393 PVAAEPATPLKAKEAT---KKKKKQFGKKRRIPT 426
DB 473 -----PLTAPHKTVSQKKRIAVVQPEKQLPPT 500

RESULT 13

ABG14398
ID ABG14398 standard; Protein; 205 AA.

XX AC ABG14398;
XX

DT 18-FEB-2002 (first entry)
XX

DE Novel human diagnostic protein #14389.
XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX

OS Homo sapiens.
XX

PN WO200175067-A2.
XX

PD 11-OCT-2001.
XX

XX 30-MAR-2001; 2001WO-US08631.
XX

XX 31-MAR-2000; 2000US-0540217.
XX

XX 23-AUG-2000; 2000US-0649167.
XX

XX (HYSE-) HYSEQ INC.
XX

XX Drmanac RT, Liu C, Tang YT;
XX

XX WPI; 2001-639362/73.
XX

XX N-PSDB; AAS78585.
XX

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX

PS Claim 20; SEQ ID NO 44757; 103pp; English.
XX

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 205 AA;

Query Match 21.9%; Score 587; DB 22; Length 205;
 Best Local Similarity 56.1%; Pred. No. 2.8e-52;
 Matches 11; Conservative 28; Mismatches 47; Indels 12; Gaps 3;
 QY 16 PSRYWIASVIQAGYVRLRYEGFENDASHDFWNCNLGTVDVHPICWAINSKILVPPPT 75
 DB 14 PSRV---ALEQECGYNALLRYEGFENDSLDFWNCNLGTVDVHPICWAINSKILVPPPT 69
 QY 76 IHAFTDKWGLMKRLVGSRTLPVDPHIKMVESMKYPRFGMRLEVVVQSVSRTRMAV 135
 DB 70 IQHNTNKAFLVRLTGAKTLPPDFPSQKVSSESNQYPRFCMRVVDVDRHLCRIKRVAV 129
 QY 136 DTVIGGLRLLYBDG-DSDDDFWCHWMSPLIHPVGSRRVGHGIMKMSRRSDMAHPTFR 194
 DB 130 ESVGGLRLVYSESDRTDDFWCHWMSPLIHPVGHGIMKMSRRSDMAHPTFR 185
 QY 195 KIYCDVAPYLFKKRVAVY 212
 DB 186 ---FDTPPHLEFAKVSFY 200

RESULT 14
 ABU11746
 ID ABU11746 standard; Protein; 538 AA.
 XX
 AC ABU11746;

13-FEB-2003 (first entry)
 DE Human MDDT polypeptide SEQ ID 693.

XX MDDT; human; disease detection and treatment molecule polypeptide;
 KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;
 KW haemostatic; nephrotropic; antianaemic; antipsoriatic; hepatotropic;
 KW gene therapy; protein replacement therapy; cell proliferative disorder;
 KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; sarcoma;
 KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;
 KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopenia;
 KW psoriasis; hepatitis.
 XX Homo sapiens.
 XX
 PN WO200279449-A2.
 XX
 PD 10-OCT-2002.
 XX
 XX 27-MAR-2002; 2002WO-US09944.
 XX
 XX 28-MAR-2001; 2001US-279619P.
 PR 29-MAR-2001; 2001US-280067P.

PR 29-MAR-2001; 2001US-280069P.
 PR 16-MAY-2001; 2001US-291280P.
 PR 17-MAY-2001; 2001US-291829P.
 PR 17-MAY-2001; 2001US-291849P.
 PR 19-JUN-2001; 2001US-299428P.
 PR 20-JUN-2001; 2001US-299776P.
 PR 20-JUN-2001; 2001US-300001P.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
 PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Ameshey SR;
 PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Klesfeld Y, Gerstin EH;
 PI Peralta CH, David KH, Lewis SA, Chen AJ, Panzer SR, Harris B;
 PI Flores V, Marwaha R, Lo A, Lan RY, Urashka MB;
 XX WPI; 2003-058431/05.
 DR N-PSDB; ABX34736.

XX New purified disease detection and treatment molecule proteins and
 PT polynucleotides, useful for diagnosing, treating or preventing cancers
 PT (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis
 PT or hepatitis
 CC Claim 27; SEQ ID NO 693; 339pp + Sequence Listing: English.

CC This invention describes a novel disease detection and treatment molecule
 CC polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,
 CC osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,
 CC antianaemic, antipsoriatic and hepatotropic activity. The polynucleotides
 CC and the polypeptides of the invention can be used for gene therapy,
 CC protein replacement therapy and are useful for treating a variety of
 CC diseases or conditions. These polypeptides or polynucleotides are
 CC particularly useful for diagnosing, treating or preventing cell
 CC proliferative disorders (e.g. cancers including adenocarcinoma,
 CC leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's
 CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's
 CC syndromes, inflammation, osteoporosis, thrombocytopenia, psoriasis or
 CC hepatitis. ABU11450-ABU11845 represent the MDDT polynucleotides encoded
 CC by ABU11450-ABU11845, described in the disclosure of the invention.
 CC NOTE: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 538 AA;

Query Match Similarity 19.0%; Score 510; DB 24; Length 538;
 Best Local Similarity 27.4%; Pred. No. 1.3e-43;
 Matches 132; Conservative 50; Mismatches 144; Indels 156; Gaps 12;

QY 3 GMKVEVLNSDAVLPSRVYVIASTVQAGYVRLRYEGFENDASHDFWNCNLGTVDVHPIGW 62
 DB 27 GMKLEGI--DPQHPS-MYFILTVAEVCYGLRLHFDGYS-CHDFVWNANSPDIHPAGW 81
 QY 63 CAINSKILVPPRTHAKFTDWKGLMKRLVGSRTLPVDPHIKMVESMKYP---PQGMRL 119
 DB 82 FEKTGHKLQPPKGYKEEFSSQYLR-----STRAQAAPKELFVSQSHPPLGFGVQMKL 137
 QY 120 EWDKSVQSVSRTRMAVDTVIGGLRLLYBDGSDDDDFWCHWMSPLIHPVGSRRVGHGIM 179
 DB 138 EAVDRMPSLCVAVSDVVDVDSRLVHFDWDDTYDWCDPSSYIHPVGW----- 188
 QY 180 MSRRSDMAHPTFRKIYCDVAPYLFKKRVAVYEGGWFEGMKLEAIDPLNLGNICVAT 239
 DB 189 ----- 188
 QY 240 VKVLLDGYLMICVDGGPSTGLDWFYHASSHAIFPATFCKNDIELTPPKGY-EAQT 298
 DB 189 -----CQKQKELTPPDYDPDNF 208
 QY 299 NWENYLEKTSKAAPSRLFNMDCPNHGFKVGMKLEAVDLMEPRICVATVKRVVHRLLSI 358
 DB 209 CWKYLEETGASAVFTWAFKVR-PPHSLVNMKLEAVDRRNPAIRVASVEDVEDHRIKI 267

QY 359 HFDGWDSEYDQWDCSPDIYFVGWCELTGYQLQPPVA-AEPAT-----PLKAKEATK 410
 DB 268 HFDGWSHGDFWIDADHDHPACWCKTGHPLOPLGPREPSSASPGGCPPLSYRSLPH 327
 QY 411 KKKQFGKKRKRIP-----PTKTRPLRQSKKPLEDDPQGA 447
 DB 328 TRTSKYSFHHKCKPTPGCGSHVTKGKTAHCLSGCPLAER--NQSRKKAELSDSEASA 385
 QY 448 RK 449
 DB 386 RK 387

RESULT 15
 ABU11833
 ID ABU11833 standard; Protein; 538 AA.
 AC ABU11833;
 XX
 DT 13-FEB-2003 (first entry)
 XX
 DE Human MDDT polypeptide SEQ ID 780.
 XX
 KW MDDT: human; disease detection and treatment molecule polypeptide;
 KW anti-inflammatory; immunosuppressive; osteoporosis; cytostatic; anti-HIV;
 KW haemostatic; nephrotropic; antianaemic; antiproliferative; hepatotropic;
 KW gene therapy; protein replacement therapy; cell proliferative disorder;
 KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; sarcoma;
 KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;
 KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopenia;
 KW psoriasis; hepatitis.
 XX
 OS Homo sapiens.
 XX
 FN WO200279449-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 27-MAR-2002; 2002WO-US09944.
 XX
 PR 28-MAR-2001; 2001US-279619P.
 PR 29-MAR-2001; 2001US-280067P.
 PR 29-MAR-2001; 2001US-280069P.
 PR 16-MAY-2001; 2001US-291280P.
 PR 17-MAY-2001; 2001US-291829P.
 PR 17-MAY-2001; 2001US-291849P.
 PR 19-JUN-2001; 2001US-299428P.
 PR 20-JUN-2001; 2001US-299776P.
 PR 20-JUN-2001; 2001US-300001P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Dafo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
 PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Anshey SR;
 PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
 PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B,
 PI Flores V, Marwaha R, Lo A, Lan RY, Urashka M;
 XX
 WIPI: 2003-058431/05.
 DR N-PSDB; ABX34823.
 XX
 XX New purified disease detection and treatment molecule proteins and
 PT polynucleotides, useful for diagnosing, treating or preventing cancers
 PT (e.g. leukemia or sarcoma), anaemia, Crohn's disease, AIDS, osteoporosis
 PT or hepatitis
 XX
 FS Claim 27; SEQ ID NO 780; 339pp + Sequence Listing; English.
 XX
 CC This invention describes a novel disease detection and treatment molecule
 CC polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,
 CC osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,
 CC antianaemic, antiproliferative and hepatotropic activity. The polynucleotides

CC and the polypeptides of the invention can be used for gene therapy,
 CC protein replacement therapy and are useful for treating a variety of
 CC diseases or conditions. These polypeptides or polynucleotides are
 CC particularly useful for diagnosing, treating or preventing cell
 CC proliferative disorders (e.g. cancers including adenocarcinoma,
 CC leukaemia, lymphoma, melanoma or sarcoma), anaemia, Crohn's
 CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's
 CC syndromes, inflammation, osteoporosis, thrombocytopenia, psoriasis or
 CC hepatitis. ABU11450-ABU11845 represent the MDDT polynucleotides encoded
 CC by ABU11450-ABU11845, described in the disclosure of the invention.
 CC NOTE: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 538 AA;

Query Match 19.0%; Score 510; DB 24; Length 538;
 Best Local Similarity 27.4%; Pred. No. 1.3e-43;
 Matches 132; Conservative 50; Mismatches 144; Indels 156; Gaps 12;

QY 3 GMKVELNSDAVLPSSRVYVIASVTQAGYRVLLYEGFENDASHDFWNLGTVDVHPGW 62
 DB 27 GMKLEGI--DPQHPS-MYFILTVAEVCYRLRLHFDGYSE--CHDFWVNANSPDIHPAGW 81
 QY 63 CAINSKILVPRTIHAKEFTDKGVMKELVGSRTLVPDFHIKMYESMKYP---FQGNRL 119
 DB 82 FEKTHKLPQPKGYKEBEFSQYLK----STRAQAAPKHLFVSQSHSPPLGFQVGMKL 137
 QY 120 EYVDKQSVSRTRMAVDVTIVIGRLRLLYEDGSDDDDFWCHMWSPLIHPVGMWSRRVHGK 179
 DB 138 EAVDRMNSLVCVASVTDVDSRFLVHFDNWDIVYWCDPSSPYIHPVGW----- 188
 QY 180 MSERSDMAHPTFRKIYCDVPIYFKKRVAYTTEGGWFEGMKLEAIDPLNLGNCVAT 239
 DB 189 ----- 188
 QY 240 VCKVLLDGYLMICVDGGSPSTGLDFWFCYHASSHAIFPATFCQKNDIELTPPKGY-EACTF 298
 DB 189 -----CQKQKPLTPPDYFDPDNF 208
 QY 299 NWNYLEKTSKAAPSRFLNMDCPNHGFKVGMKLEAVDLMEPRLICVATVXRVVHRLLSI 358
 DB 209 CWKYLEETGASAVPTWAFKVR-PPHSFLVNMKLEAVDRRNPALIRVASVEDVEDHRIKI 267
 QY 359 HFDGWDSEYDQWDCSPDIYFVGWCELTGYQLQPPVA-AEPAT-----PLKAKEATK 410
 DB 268 HFDGWSHGDFWIDADHDHPACWCKTGHPLOPLGPREPSSASPGGCPPLSYRSLPH 327
 QY 411 KKKQFGKKRKRIP-----PTKTRPLRQSKKPLEDDPQGA 447
 DB 328 TRTSKYSFHHKCKPTPGCGSHVTKGKTAHCLSGCPLAER--NQSRKKAELSDSEASA 385
 QY 448 RK 449
 DB 386 RK 387

Search completed: February 4, 2004, 09:06:54
 Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2004, 09:05:51 : Search time 21 Seconds
(without alignments)
997.327 Million cell updates/sec

Title: US-10-031-915-36

Perfect score: 2683

Sequence: 1 MKGMKEVLNSDAVLPSRV.....KASSPELPVSVNIKETDD 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/6C COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance, to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	357.5	13.3	591	2	US-08-852-153-6
2	357.5	13.3	620	2	US-08-852-153-2
3	353.5	13.2	577	2	US-08-852-153-4
4	349	13.0	664	2	US-08-852-153-8
5	97.5	3.6	739	1	US-07-803-6225-2
6	91.5	3.4	925	4	US-09-116-473-2
7	90.5	3.4	425	2	US-08-732-028-2
8	90.5	3.4	425	4	US-09-183-228-2
9	90	3.4	339	2	US-08-834-655-4
10	90	3.4	339	3	US-08-834-033A-4
11	90	3.4	339	4	US-09-363-574-4
12	90	3.4	339	4	US-09-363-526-4
13	90	3.4	339	4	US-09-330-235-20
14	90	3.4	623	4	US-08-714-741-47
15	89.5	3.3	658	4	US-09-252-991A-26319
16	89	3.3	1022	4	US-09-344-510B-1
17	88.5	3.3	901	3	US-08-836-135-22
18	88.5	3.3	906	3	US-08-836-135-24
19	88.5	3.3	903	3	US-08-836-135-8
20	88.5	3.3	909	3	US-08-836-135-10
21	88.5	3.3	914	3	US-08-836-135-12
22	88.5	3.3	926	3	US-08-836-135-14
23	88.5	3.3	931	3	US-08-836-135-16
24	88	3.3	486	2	US-08-942-423-3
25	87.5	3.3	425	3	US-09-096-776B-8
26	87.5	3.3	425	4	US-09-923-922-8
27	87.5	3.3	436	3	US-08-840-767-6

28	87.5	3.3	614	4	US-09-328-352-5476	Sequence 5476, Ap
29	87	3.2	1626	2	US-08-771-602D-2	Sequence 2, Appli
30	87	3.2	1626	2	US-09-232-446B-2	Sequence 2, Appli
31	86	3.2	231	4	US-09-252-991A-26289	Sequence 26289, A
32	86	3.2	935	4	US-09-271-438A-3	Sequence 3, Appli
33	86	3.2	935	4	US-09-271-438A-8	Sequence 8, Appli
34	86	3.2	1440	3	US-09-357-251-37	Sequence 37, Appli
35	86	3.2	1512	4	US-09-443-184-48	Sequence 48, Appli
36	85.5	3.2	976	3	US-08-894-997-50	Sequence 50, Appli
37	85	3.2	708	1	US-08-396-479B-8	Sequence 8, Appli
38	85	3.2	708	1	US-08-818-823-8	Sequence 10, Appli
39	85	3.2	739	1	US-08-396-479B-10	Sequence 10, Appli
40	85	3.2	739	1	US-08-818-823-10	Sequence 10, Appli
41	85	3.2	1068	1	US-08-396-479B-12	Sequence 12, Appli
42	85	3.2	1068	1	US-08-818-823-12	Sequence 12, Appli
43	85	3.2	1075	5	PCT-US94-07297-41	Sequence 41, Appli
44	84.5	3.1	1105	3	US-08-999-774A-2	Sequence 2, Appli
45	84	3.1	184	4	US-09-252-991A-27862	Sequence 27862, A

ALIGNMENTS

RESULT 1
US-08-852-153-6
; Sequence 6, Application US/08852153
; Patent No. 5914366
; GENERAL INFORMATION:
; APPLICANT: Randazzo, Filippo
; TITLE OF INVENTION: Mammalian Sex Comb on Midleg Acts as a Tumor Suppressor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,153
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Guth, Joseph H.
; REGISTRATION NUMBER: 31,261
; REFERENCE/DOCKET NUMBER: 1224.006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-3888
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 591 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-852-153-6

Query Match 13.3%; Score 357.5; DB 2; Length 591;

Best Local Similarity 35.9%; Pred. No. 4.2e-31;

Matches 102; Conservative 31; Mismatches 102; Indels 49; Gaps 11;

QY 218 FEQMKLEADPLNLGNICVATYCKVLLDGYLMICVDGSPSTGLDWFCHASHAIFPA 277

DB 11 FKSMKLEADPPRTTSTCIATVVG-LTGARLRLDG----SDKNDFWFLVDSAEIQPI 66

QY 278 TFQCKNDIELTPPKGYEAQFFNENYLEKTKSKA--APSLFNMD--CPNHG-FKVGMLK 332

Db 67 GNCENKGMQLPPLGFRNLNASSWPMFLKLTNLGAEMAPIRI FHKBPSPSHNFFKMGKML 126
Qy 333 EAVDLMEPRLICVATKRVVHRLLSIHFDGWDSEYDQWDCESPDYFVGMWCELGYQLQ 392
Db 127 EAVDRKNPHFICPATIGEVRGSEVLVTFDWRGAFDYWCFRDSRDIFFVGMWCSLTGDNILQ 186
Qy 393 PP--VAEPATPLKAEATKKK-----KQFGKKRKRIIP----- 424
Db 187 PPQTKVVPKPNYPASDVNTEKPSIHSSTKTVLEHQPQGRKPKGKGRTPKTLISHPI 246
Qy 425 --PTKT-RPL-----RQSK-KPILLEDDPOGARKISSEVPV 456
Db 247 SAPSKTAEPKPKGPKGSKRKPRTLNPPASPPTTSTPEP 290

RESULT 2
US-08-852-153-2
; Sequence 2, Application US/08852153
; Patent No. 5914266
; GENERAL INFORMATION:
; APPLICANT: Randazzo, Filippo
; TITLE OF INVENTION: Mammalian Sex Comb on Midleg Acts as a Tumor Suppressor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,153
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Guth, Joseph H.
; REGISTRATION NUMBER: 31,261
; REFERENCE/DOCKET NUMBER: 1224.006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-3888
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 620 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-852-153-2

Query Match 13.3%; Score 357.5; DB 2; Length 620;
Best Local Similarity 35.9%; Pred. No. 4.5e-31;
Matches 102; Conservative 31; Mismatches 102; Indels 49; Gaps 11;

Qy 218 FEEGMKLEADPLNLGNICVATVKVLLDGYLMICVDGPGSTGLDWFVCHASSHAIFPA 277
Db 40 FKISMKLEAQDPRNTTSTCIATVVG-LTGARLRLDG---SDNKDFWRLVDSAEIQPI 95
Qy 278 TFCQNDIELTPPKGYEAQTFNWNYLEKTKSKA--APSRLENMD--CPNHG-FKVGKML 332
Db 96 GNCENKGMQLPPLGFRNLNASSWPMFLKLTNLGAEMAPIRI FHKBPSPSHNFFKMGKML 155
Qy 333 EAVDLMEPRLICVATKRVVHRLLSIHFDGWDSEYDQWDCESPDYFVGMWCELGYQLQ 392
Db 156 EAVDRKNPHFICPATIGEVRGSEVLVTFDWRGAFDYWCFRDSRDIFFVGMWCSLTGDNILQ 215
Qy 393 PP--VAEPATPLKAEATKKK-----KQFGKKRKRIIP----- 424

Db 216 PPQTKVVPKPNYPASDVNTEKPSIHSSTKTVLEHQPQGRKPKGKGRTPKTLISHPI 275
Qy 425 --PTKT-RPL-----RQSK-KPILLEDDPOGARKISSEVPV 456
Db 276 SAPSKTAEPKPKGPKGSKRKPRTLNPPASPPTTSTPEP 319

RESULT 3
US-08-852-153-4
; Sequence 4, Application US/08852153
; Patent No. 5914266
; GENERAL INFORMATION:
; APPLICANT: Randazzo, Filippo
; TITLE OF INVENTION: Mammalian Sex Comb on Midleg Acts as a Tumor Suppressor
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,153
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Guth, Joseph H.
; REGISTRATION NUMBER: 31,261
; REFERENCE/DOCKET NUMBER: 1224.006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-3888
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 577 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-852-153-4

Query Match 13.2%; Score 353.5; DB 2; Length 577;
Best Local Similarity 36.1%; Pred. No. 1.1e-30;
Matches 101; Conservative 30; Mismatches 100; Indels 49; Gaps 11;

Qy 222 MKLEAIDPLNLGNICVATVKVLLDGYLMICVDGPGSTGLDWFVCHASSHAIFPAFCQ 281
Db 1 MKLEAQDPRNTTSTCIATVVG-LTGARLRLDG---SDNKDFWRLVDSAEIQPENCE 56
Qy 282 KNDIELTPPKGYEAQTFNWNYLEKTKSKA--APSRLENMD--CPNHG-FKVGKMLEAVD 336
Db 57 KNGGMQLPPLGFRNLNASSWPMFLKLTNLGAEMAPIRI FHKBPSPSHNFFKMGKMLEAVD 116
Qy 337 LMEPRLICVATKRVVHRLLSIHFDGWDSEYDQWDCESPDYFVGMWCELGYQLQPP-- 394
Db 117 RKNPHFICPATIGEVRGSEVLVTFDWRGAFDYWCFRDSRDIFFVGMWCSLTGDNILQPPGT 176
Qy 395 VAEPATPLKAEATKKK-----KQFGKKRKRIIP-----PT 426
Db 177 KVVIPKPNYPASDVNTEKPSIHSSTKTVLEHQPQGRKPKGKGRTPKTLISHPI SAPS 236
Qy 427 KT-RPL-----RQSK-KPILLEDDPOGARKISSEVPV 456
Db 237 KTAEPLPKPKGPKGPKGSKRKPRTLNPPASPPTTSTPEP 276

RESULT 4

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2004, 09:08:22 ; Search time 39 Seconds
(without alignments)
2657.544 Million cell updates/sec

Title: US-10-031-915-36

Perfect score: 2683

Sequence: 1 MKGMKEVLNSDAVLPSRV.....KASSPELPVSNIKQETDD 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2679	99.9	705	12	US-10-094-749-2556
2	2185	81.4	526	12	US-10-108-260A-4101
3	1688	62.9	329	12	US-10-264-237-1824
4	1410	52.6	631	9	US-09-789-919-56
5	1387	51.7	410	10	US-09-872-523-68
6	495.5	18.5	623	12	US-10-104-047-3512
7	494.5	18.4	498	10	US-09-872-523-5
8	493.5	18.4	498	10	US-09-872-523-71
9	486.5	18.1	498	10	US-09-872-523-72
10	440	16.4	78	12	US-10-029-386-27484
11	417	15.5	77	9	US-09-864-761-38826
12	417	15.5	77	12	US-10-029-386-30438
13	398	14.8	427	12	US-10-104-047-2262
14	389.5	14.5	406	12	US-10-104-047-3204
15	373.5	13.9	217	12	US-10-264-049-2539

16	340	12.7	827	9	US-09-801-574-12
17	330.5	12.3	100	10	US-09-872-523-7
18	302.5	11.3	411	10	US-09-872-523-70
19	281.5	10.5	125	15	US-10-204-887-93
20	272	10.1	512	10	US-09-872-523-69
21	271.5	10.1	96	10	US-09-872-523-10
22	268.5	10.0	100	10	US-09-872-523-9
23	259.5	9.7	186	12	US-10-104-047-2159
24	242.5	9.0	311	10	US-09-764-868-1009
25	238.5	8.9	100	10	US-09-872-523-12
26	236.5	8.8	100	10	US-09-872-523-8
27	233.5	8.3	100	10	US-09-872-523-15
28	211	7.9	99	10	US-09-872-523-11
29	199	7.4	38	9	US-09-864-761-38862
30	186.5	7.0	96	10	US-09-872-523-13
31	178	6.6	99	10	US-09-872-523-14
32	177	6.6	67	9	US-09-864-761-36870
33	162	6.0	78	9	US-09-864-761-48140
34	158	5.9	74	12	US-10-029-386-33605
35	141	5.3	25	9	US-09-864-761-42641
36	122	4.5	25	9	US-09-864-761-38882
37	111	4.1	790	12	US-10-369-493-3803
38	102	3.8	512	12	US-10-320-797-3205
39	94	3.5	634	12	US-10-369-493-13304
40	93.5	3.5	371	9	US-09-895-943-5
41	93.5	3.5	371	10	US-09-895-593-5
42	93.5	3.5	371	11	US-09-376-430-2
43	93.5	3.5	371	12	US-10-354-951-3
44	93.5	3.5	371	14	US-10-008-566-4
45	93.5	3.5	371	14	US-10-078-059-2

ALIGNMENTS

RESULT 1
US-10-094-749-2556
; Sequence 2556, Application US/10094749
; Publication No. US20030219741A1

GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NACHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2556
; LENGTH: 705
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2556

Sequence 12, Appl
Sequence 7, Appl
Sequence 70, Appl
Sequence 93, Appl
Sequence 69, Appl
Sequence 10, Appl
Sequence 9, Appl
Sequence 2159, Ap
Sequence 1009, Ap
Sequence 12, Appl
Sequence 8, Appl
Sequence 15, Appl
Sequence 11, Appl
Sequence 38862, A
Sequence 13, Appl
Sequence 14, Appl
Sequence 36870, A
Sequence 48140, A
Sequence 33605, A
Sequence 42641, A
Sequence 38882, A
Sequence 3803, Ap
Sequence 3205, Ap
Sequence 13304, A
Sequence 5, Appl
Sequence 2, Appl
Sequence 3, Appl
Sequence 4, Appl
Sequence 2, Appl

Query Match 99.9%; Score 2679; DB 12; Length 705;
Best Local Similarity 99.8%; Pred. No. 4.5e-259;
Matches 494; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKGKVEVLNSDAVLPSRVVWIASVIQTAGYRVLLRYEGFENDASHDFWNLGTVDVHPI 60
DB 211 MKGKVEVLNSDAVLPSRVVWIASVIQTAGYRVLLRYEGFENDASHDFWNLGTVDVHPI 270
QY 61 GWCAINSKILVPPRTTHAKFTDMKGYLMKELVGSRTLPVDFHFKWVESMKYPFQGMRL 120
DB 271 GWCAINSKILVPPRTTHAKFTDMKGYLMKELVGSRTLPVDFHFKWVESMKYPFQGMRL 330
QY 121 VVDKSVSRTRMAVVDTVIGRLRLLYEDGSDDDDFWCHWMSPLIHPVGSRRVGHGIM 180
DB 331 VVDKSVSRTRMAVVDTVIGRLRLLYEDGSDDDDFWCHWMSPLIHPVGSRRVGHGIM 390
QY 181 SERRSDMAHPTFRKIYCDVAVPYLFKKVRAVYEGGFEGMKLEAIDPLNLGNICVATV 240
DB 391 SERRSDMAHPTFRKIYCDVAVPYLFKKVRAVYEGGFEGMKLEAIDPLNLGNICVATV 450
QY 241 CKVLLDGYLMICVDGGSTDGLDFWFCYHASSHAIFPATFCQKNDIELTTPKGYEAQTFNW 300
DB 451 CKVLLDGYLMICVDGGSTDGLDFWFCYHASSHAIFPATFCQKNDIELTTPKGYEAQTFNW 510
QY 301 ENYLEKTSKAAPSRFLNMDCPNHGFKVGMKLEAVDLMEPRLICVATVKRVVHRLLSIH 360
DB 511 ENYLEKTSKAAPSRFLNMDCPNHGFKVGMKLEAVDLMEPRLICVATVKRVVHRLLSIH 570
QY 361 DGMSEYDQWDCSPDIYPVGMCELTYQLQPPVAAEPATPLFAEATKTKKKQFGKKR 420
DB 571 DGMSEYDQWDCSPDIYPVGMCELTYQLQPPVAAEPATPLFAEATKTKKKQFGKKR 630
QY 421 KRIPPTKTRPLROGSKPILLEDPOGARKISSEPVFGEIIAVRVKSEHLDVASPKKASP 480
DB 631 KRIPPTKTRPLROGSKPILLEDPOGARKISSEPVFGEIIAVRVKSEHLDVASPKKASP 690
QY 481 ELPVSVENIKQETDD 495
DB 691 ELPVSVENIKQETDD 705

RESULT 2
US-10-108-260A-4101
; Sequence 4101, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cdna
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; PRIOR FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 4101
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-4101

Query Match 81.4%; Score 2185; DB 12; Length 526;
Best Local Similarity 100.0%; Pred. No. 8.9e-210;
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKGKVEVLNSDAVLPSRVVWIASVIQTAGYRVLLRYEGFENDASHDFWNLGTVDVHPI 60
DB 120 MKGKVEVLNSDAVLPSRVVWIASVIQTAGYRVLLRYEGFENDASHDFWNLGTVDVHPI 179
QY 61 GWCAINSKILVPPRTTHAKFTDMKGYLMKELVGSRTLPVDFHFKWVESMKYPFQGMRL 120
DB 180 GWCAINSKILVPPRTTHAKFTDMKGYLMKELVGSRTLPVDFHFKWVESMKYPFQGMRL 239
QY 121 VVDKSVSRTRMAVVDTVIGRLRLLYEDGSDDDDFWCHWMSPLIHPVGSRRVGHGIM 180

DB 240 VVDKSVSRTRMAVVDTVIGRLRLLYEDGSDDDDFWCHWMSPLIHPVGSRRVGHGIM 299
QY 181 SERRSDMAHPTFRKIYCDVAVPYLFKKVRAVYEGGFEGMKLEAIDPLNLGNICVATV 240
DB 300 SERRSDMAHPTFRKIYCDVAVPYLFKKVRAVYEGGFEGMKLEAIDPLNLGNICVATV 359
QY 241 CKVLLDGYLMICVDGGSTDGLDFWFCYHASSHAIFPATFCQKNDIELTTPKGYEAQTFNW 300
DB 360 CKVLLDGYLMICVDGGSTDGLDFWFCYHASSHAIFPATFCQKNDIELTTPKGYEAQTFNW 419
QY 301 ENYLEKTSKAAPSRFLNMDCPNHGFKVGMKLEAVDLMEPRLICVATVKRVVHRLLSIH 360
DB 420 ENYLEKTSKAAPSRFLNMDCPNHGFKVGMKLEAVDLMEPRLICVATVKRVVHRLLSIH 479
QY 361 DGMSEYDQWDCSPDIYPVGMCELTYQLQPPVAA 397
DB 480 DGMSEYDQWDCSPDIYPVGMCELTYQLQPPVAA 516

RESULT 3
US-10-264-237-1824
; Sequence 1824, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PAL31P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; PRIOR FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: Patent in Ver. 3.1
; SEQ ID NO 1824
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: MISC FEATURE
; NAME/KEY: MISC FEATURE
; LOCATION: (185)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (223)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-237-1824

Query Match 62.9%; Score 1688; DB 12; Length 329;
Best Local Similarity 99.0%; Pred. No. 2.8e-160;
Matches 308; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MKGKVEVLNSDAVLPSRVVWIASVIQTAGYRVLLRYEGFENDASHDFWNLGTVDVHPI 60
DB 1 MKGKVEVLNSDAVLPSRVVWIASVIQTAGYRVLLRYEGFENDASHDFWNLGTVDVHPI 60
QY 61 GWCAINSKILVPPRTTHAKFTDMKGYLMKELVGSRTLPVDFHFKWVESMKYPFQGMRL 120
DB 61 GWCAINSKILVPPRTTHAKFTDMKGYLMKELVGSRTLPVDFHFKWVESMKYPFQGMRL 120
QY 121 VVDKSVSRTRMAVVDTVIGRLRLLYEDGSDDDDFWCHWMSPLIHPVGSRRVGHGIM 180
DB 121 VVDKSVSRTRMAVVDTVIGRLRLLYEDGSDDDDFWCHWMSPLIHPVGSRRVGHGIM 180
QY 181 SERRSDMAHPTFRKIYCDVAVPYLFKKVRAVYEGGFEGMKLEAIDPLNLGNICVATV 240
DB 181 SERRSDMAHPTFRKIYCDVAVPYLFKKVRAVYEGGFEGMKLEAIDPLNLGNICVATV 240
QY 241 CKVLLDGYLMICVDGGSTDGLDFWFCYHASSHAIFPATFCQKNDIELTTPKGYEAQTFNW 300
DB 241 CKVLLDGYLMICVDGGSTDGLDFWFCYHASSHAIFPATFCQKNDIELTTPKGYEAQTFNW 300

```
QY 301 ENYLEKTSKA 311
|||||
Db 301 ENYLEKTSKA 311

RESULT 4
US-09-789-919-56
; Sequence 56, Application US/09789919
; Patent No. US20020064855A1
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ibor
; APPLICANT: Moore, Kateri
; TITLE OF INVENTION: GENES THAT REGULATE HEMATOPOIETIC BLOOD FORMING STEM
; TITLE OF INVENTION: CELLS AND USES THEREOF
; FILE REFERENCE: 2275-1-005
; CURRENT APPLICATION NUMBER: US/09/789,919
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 56
; LENGTH: 631
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-789-919-56

Query Match 52.6%; Score 1410; DB 9; Length 631;
Best Local Similarity 56.2%; Pred. No. 5.1e-132;
Matches 263; Conservative 67; Mismatches 122; Indels 16; Gaps 4;

QY 2 KGMKEVLSNDAVLPSRVYTIASVIQTAGYRVLLRYEGFENDASHDFWNCNLGTVDVHPG 61
|||||
Db 177 ENVAIEVNTDCSLPTKVFVIAIGIILAGYNALLRYEGFENDSLDFWNCNICSDIHPVG 236

QY 62 WCAINSKILVPPRTIHAFTDKWGYLMKRLVGSRTLPVDPHIKVSMKYPPFQGMRLV 121
|||||
Db 237 WCAASGKPLVPPRTVQHKYTNWKAFLVKRLTGAKTLPPDFSQKVSSEMQYPPKCMRVEV 296

QY 122 VDKSQVSRMAVVDVTIGRRLILYEDG-DSDDDFCHMWSPLIHPVGNRSRVGHGKIM 180
|||||
Db 297 VDKHLCRTVAVVESVIGRRLVIEESDGDGDFWCHMWSPLIHHIGMSRSGHFRFR 356

QY 181 SERRSDMAHPTFRKIYCDAPVLFKKVRAVYTEGGWFEEGKMLEAIDPLNLGNICVATV 240
|||||
Db 357 SDITPKQDGH-----FDTPPHLFAKVEVDQSGEWFEGKMLEAIDPLNLSTICVATI 409

QY 241 CKVLLDGLYLMICVDGSPSTGLDFWCFHASSHAIFPATPCOKNDIELTPPKGYEATFNW 300
|||||
Db 410 RKVLADGFLMIGIDGSEADGSDWFCYHATSPSIFPVGFCFCEINMIELTPPRGYTKLPFKW 469

QY 301 ENYLEKTSKAAPSRFLNMDCPNHGFKVGMKLEAVDLMEPRLICVATVKRVHRLLSIHF 360
|||||
Db 470 FDIYLETGSAAPVKLFNKDVPNHGFRVGMKLEAVDLMEPRLICVATVTRIHLRLIHF 529

QY 361 DGMSEYDQWDCSPDIYPVGMCELTGYQLQPPVAAEATPLKAKATKK-KKQFGKK 419
|||||
Db 530 DGMSEYDQWDCSPDLNPNVGMCELTGYQLQPPASQSSRESQSSAKSKAKAKSQYKG 589

QY 420 RKIPPTKTRPLQSGKKPLEDDPGARKISSEPVGPIIIVRVKEE 467
|||||
Db 590 HKKMTTSQ-----KEELLGDGYSFLHGASDQSSNGSATVYIKQE 630

RESULT 5
US-09-782-523-68
; Sequence 68, Application US/09872523
; Patent No. US20020137906A1
; GENERAL INFORMATION:
; APPLICANT: Horvitz, H. Robert
; APPLICANT: Davison, Ewa M.
; APPLICANT: Lu, Xiaowei
; TITLE OF INVENTION: A Tumor Suppressor Pathway in C. Elegans
; FILE REFERENCE: 01997/536002
```

```
; CURRENT APPLICATION NUMBER: US/09/872,523
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: US 60/208,802
; PRIOR FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-872-523-68

Query Match 51.7%; Score 1387; DB 10; Length 410;
Best Local Similarity 62.6%; Pred. No. 5.4e-130;
Matches 248; Conservative 56; Mismatches 84; Indels 8; Gaps 2;

QY 2 KGMKEVLSNDAVLPSRVYTIASVIQTAGYRVLLRYEGFENDASHDFWNCNLGTVDVHPG 61
|||||
Db 10 ENVRVEVNTDCSLPTKVFVIAIGIILAGYNALLRYEGFENDSLDFWNCNICSDIHPVG 69

QY 62 WCAINSKILVPPRTIHAFTDKWGYLMKRLVGSRTLPVDPHIKVSMKYPPFQGMRLV 121
|||||
Db 70 WCAASGKPLVPPRTIHAFTDKWGYLMKRLVGSRTLPVDPHIKVSMKYPPFQGMRLV 129

QY 122 VDKSQVSRMAVVDVTIGRRLILYEDG-DSDDDFCHMWSPLIHPVGNRSRVGHGKIM 180
|||||
Db 130 VDKHLCRTVAVVESVIGRRLVIEESDGDGDFWCHMWSPLIHHIGMSRSGHFRFR 189

QY 181 SERRSDMAHPTFRKIYCDAPVLFKKVRAVYTEGGWFEEGKMLEAIDPLNLGNICVATV 240
|||||
Db 190 SDITPKQDGH-----DPPHLFAKVEVDQSGEWFEGKMLEAIDPLNLSTICVATI 242

QY 241 CKVLLDGLYLMICVDGSPSTGLDFWCFHASSHAIFPATPCOKNDIELTPPKGYEATFNW 300
|||||
Db 243 RKVLADGFLMIGIDGSEADGSDWFCYHATSPSIFPVGFCFCEINMIELTPPRGYTKLPFKW 302

QY 301 ENYLEKTSKAAPSRFLNMDCPNHGFKVGMKLEAVDLMEPRLICVATVKRVHRLLSIHF 360
|||||
Db 303 FDIYLETGSAAPVKLFNKDVPNHGFRVGMKLEAVDLMEPRLICVATVTRIHLRLIHF 362

QY 361 DGMSEYDQWDCSPDIYPVGMCELTGYQLQPPVAA 396
|||||
Db 363 DGMSEYDQWDCSPDLNPNVGMCELTGYQLQPPAS 398

RESULT 6
US-10-104-047-3512
; Sequence 3512, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3512
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3512

Query Match 18.5%; Score 495.5; DB 12; Length 623;
Best Local Similarity 36.6%; Pred. No. 2.1e-40;
Matches 121; Conservative 53; Mismatches 118; Indels 39; Gaps 12;

QY 83 WGYLMKRLVGSRTLPVDPHIKVSMKYPP-----FQGMRLVWVKSQVSRTEMAVVD 137
|||||
Db 54 WEWYKSEQ--KAVAAPVELFSK---DQSPFEHENGFOIGNRLEGIDPRHFSVFCVLSVAE 108
```

Qy 138 VIGRLRLLYEDGSDSDDFWCHWMSPLIHPVGRSRRVGHGIMKSRERRSDMAHPT-PRK- 195
Db 109 VCGYRLRLHFDGYSCLVDFWNTAGSDIHPVGCERTKHEL-----HIPKGYRKD 158
Qy 196 --IYCD-----AVPYLKKVRAVYVTEGGWFECKMKELEIDPLNLGNICVATVCKVL 244
Db 159 KFWMDYLKACKLQNAKPKLFRNRSNGPMSEKFEQVGMKLEAVDRKNPSLVCAVATIADIV 218
Qy 245 LDGALMICVDGSGPSTDGLDWFCHYHASSHAIFPATFCQKNDIELTPPKGY-EAQTENWENY 303
Db 219 ED-RLLVHFDNW--DSYDYM-C-DVNSPYQPVGWCQENGRTLLAQGYPENESWTEY 274
Qy 304 LEKTKKAAPSRLEPNMDCNHSKGVMKLEAVDLMEPRLICVATKRVVHRLLSLHFDGW 363
Db 275 LEATQNAVPAKVFCKRLP-HGFLPNMKLEVDKRNPRILRVATIVDQDQKRVHFDGW 333
Qy 364 DSEYDOWDCESPDYVPGWCELTGYOLOPP 394
Db 334 DHKYDTWVADSPIHPIGNCVDTGHPLEVP 364

RESULT 7

US-09-872-523-5

; Sequence 5, Application US/09872523

; Patent No. US20020137906A1

; GENERAL INFORMATION:

; APPLICANT: Horvitz, H. Robert

; APPLICANT: Davison, Ewa M.

; APPLICANT: Lu, Xiaowei

; TITLE OF INVENTION: A Tumor Suppressor Pathway in C. Elegans

; FILE REFERENCE: 01997/536002

; CURRENT APPLICATION NUMBER: US/09/872,523

; CURRENT FILING DATE: 2001-06-01

; PRIOR APPLICATION NUMBER: US 60/208,802

; PRIOR FILING DATE: 2000-06-02

; NUMBER OF SEQ ID NOS: 78

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5

; LENGTH: 498

; TYPE: PRT

; ORGANISM: Caenorhabditis elegans

US-09-872-523-5

Query Match 18.4%; Score 494.5; DB 10; Length 498;

Best Local Similarity 29.7%; Pred. No. 1.9e-40;

Matches 132; Conservative 81; Mismatches 149; Indels 83; Gaps 18;

Qy 10 NSDAVLPSRVYVVIASVIQTAGYRVLLRYEGFENDASHDFWCNLTGVDPVHPICWCAINSK- 68

Db 72 NCDSI---QVRWPARIEKVCGRVLAQFIG---ADTKFWLNLSDDMFGLANAAMSDPN 124

Qy 69 ----ILVPRTHAKF-TWKGYLMK----RLVGSRTLPVDHFKMVESMKYPPFGMRL 119

Db 125 MDKIVVAPPLAINEEYQNDMVYNNCIDGEIVGQTSLSPKFDEGKALLSKHRFKVGQRL 184

Qy 120 EVVDKQSVSTRNAVVDVTVIGRLR--LLYED-----GSDDDD-----FWCHMWSP 163

Db 185 ELLNYSNSTEIRVARIQEICGRMNVSIITKDDPESLPDADDDRVQFSSGSQYWIDEGSF 244

Qy 164 LIHPVGM-----SRVGHGIMKSRERSDMAHPTFRKIYCDV-PYLF 205

Db 245 FIFPVGFAAVNGVQLNAKKEYIEHTNKIAQAIKNGENPRYDSDVTFDQAKDPIDPMIW 304

Qy 206 KKVRAVYTEGGWFECKMKELEIDPL--NLGNICVATV--CKVLLDGYLMICVDGSGPSTD 260

Db 305 RKVKV-----GQKPELIDPLAQQFNHLVASILKFKCT--EGYLIVGMDGPDAL 352

Qy 261 GLDWFCHYHASSHAIFPATFCQKNDIELTPPKGYEACTFNWENYLEKTKSKAAPSRLFNMD 320

Db 353 --DSFPIHINNTFMFPVGYABKYNLELVPPDFBK-TFRWDEYLEKESAEITLPLDFK-P 408

Qy 321 CPNH-----GFKVGMKLEAVDLMEPRLICVATKRVVHRLLSLHFDGWSEYDQ 369

Db 409 MPQERLDKFKVLLISKRVGLRLEAADMCENQFICPATVKSVHGRLLNVNFDGWDEEFDE 468
Qy 370 WDCESPDYVPGWCELTGYOLOPP 394
Db 469 LYVDVSHDILPIGWCEAHSYVLOPP 493

RESULT 8

US-09-872-523-71

; Sequence 71, Application US/09872523

; Patent No. US20020137906A1

; GENERAL INFORMATION:

; APPLICANT: Horvitz, H. Robert

; APPLICANT: Davison, Ewa M.

; APPLICANT: Lu, Xiaowei

; TITLE OF INVENTION: A Tumor Suppressor Pathway in C. Elegans

; FILE REFERENCE: 01997/536002

; CURRENT APPLICATION NUMBER: US/09/872,523

; CURRENT FILING DATE: 2001-06-01

; PRIOR APPLICATION NUMBER: US 60/208,802

; PRIOR FILING DATE: 2000-06-02

; NUMBER OF SEQ ID NOS: 78

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 71

; LENGTH: 498

; TYPE: PRT

; ORGANISM: Caenorhabditis elegans

US-09-872-523-71

Query Match 18.4%; Score 493.5; DB 10; Length 498;

Best Local Similarity 29.7%; Pred. No. 2.4e-40;

Matches 132; Conservative 81; Mismatches 149; Indels 83; Gaps 18;

Qy 10 NSDAVLPSRVYVVIASVIQTAGYRVLLRYEGFENDASHDFWCNLTGVDPVHPICWCAINSK- 68

Db 72 NCDSI---QVRWPARIEKVCGRVLAQFIG---ADTKFWLNLSDDMFGLANAAMSDPN 124

Qy 69 ----ILVPRTHAKF-TWKGYLMK----RLVGSRTLPVDHFKMVESMKYPPFGMRL 119

Db 125 MDKIVVAPPLAINEEYQNDMVYNNCIDGEIVGQTSLSPKFDEGKALLSKHRFKVGQRL 184

Qy 120 EVVDKQSVSTRNAVVDVTVIGRLR--LLYED-----GSDDDD-----FWCHMWSP 163

Db 185 ELLNYSNSTEIRVARIQEICGRMNVSIITKDDPESLPDADDDRVQFSSGSQYWIDEGSF 244

Qy 164 LIHPVGM-----SRVGHGIMKSRERSDMAHPTFRKIYCDV-PYLF 205

Db 245 FIFPVGFAAVNGVQLNAKKEYIEHTNKIAQAIKNGENPRYDSDVTFDQAKDPIDPMIW 304

Qy 206 KKVRAVYTEGGWFECKMKELEIDPL--NLGNICVATV--CKVLLDGYLMICVDGSGPSTD 260

Db 305 RKVKV-----GQKPELIDPLAQQFNHLVASILKFKCT--EGYLIVGMDGPDAL 352

Qy 261 GLDWFCHYHASSHAIFPATFCQKNDIELTPPKGYEACTFNWENYLEKTKSKAAPSRLFNMD 320

Db 353 --DNFPIHINNTFMFPVGYABKYNLELVPPDFBK-TFRWDEYLEKESAEITLPLDFK-P 408

Qy 321 CPNH-----GFKVGMKLEAVDLMEPRLICVATKRVVHRLLSLHFDGWSEYDQ 369

RESULT 9

US-09-872-523-72

; Sequence 72, Application US/09872523

; Patent No. US20020137906A1

; GENERAL INFORMATION:

; APPLICANT: Horvitz, H. Robert

; APPLICANT: Davison, Ewa M.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2004, 09:07:27 ; Search time 23 Seconds
(without alignments)
1675.070 Million cell updates/sec

Title: US-10-031-915-36

Perfect score: 2683

Sequence: 1 MKGMKEVLNSDAVLPSRVY.....KASSPELPVSVENIKQETDD 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 376054 seqs, 77831530 residues

Total number of hits satisfying chosen parameters: 376054

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents AA New:*

- 1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1111	41.4	891	5	US-09-614-150A-38841
2	1111	41.4	1243	5	US-09-614-150A-27069
3	611	22.8	894	6	US-10-719-993-647
4	510	19.0	538	6	US-10-473-040-693
5	510	19.0	538	6	US-10-473-040-780
6	402	15.0	1477	5	US-09-614-150A-12672
7	362.5	13.5	269	6	US-10-473-040-724
8	359	13.4	246	6	US-10-473-040-550
9	327	12.2	877	5	US-09-614-150A-20184
10	109.5	4.1	1026	7	US-60-490-890-1851
11	104	3.9	6669	1	PCT-US03-21379-80
12	104	3.9	6669	1	PCT-US03-38193-25
13	104	3.9	6669	5	US-10-723-860-25
14	102.5	3.8	779	5	US-09-614-150A-31704
15	100.5	3.7	566	6	US-10-425-114A-67814
16	100.5	3.7	574	6	US-10-425-114A-65213
17	98.5	3.7	600	6	US-10-739-930-7654
18	96.5	3.6	1377	7	US-60-490-890-2525
19	96	3.6	1065	6	US-10-648-512-42
20	95.5	3.6	1231	5	US-09-614-150A-6006
21	94	3.5	510	6	US-10-425-114A-49488
22	94	3.5	898	6	US-10-648-512-24
23	94	3.5	906	6	US-10-648-512-36
24	94	3.5	906	6	US-10-648-512-38
25	94	3.5	970	6	US-10-648-512-34
26	94	3.5	1001	6	US-10-648-512-40

27 94 3.5 1065 6 US-10-648-512-22 Sequence 22, Appl
28 94 3.5 1065 6 US-10-648-512-32 Sequence 32, Appl
29 93 3.5 414 5 US-09-614-150A-4752 Sequence 4752, Ap
30 93 3.5 626 6 US-10-679-063-17795 Sequence 17795, A
31 92.5 3.4 944 6 US-10-425-114A-55600 Sequence 55600, A
32 92.5 3.4 1578 7 US-60-495-114-1487 Sequence 1487, Ap
33 92.5 3.4 1593 1 PCT-US03-30720-855 Sequence 855, App
34 92.5 3.4 1593 1 PCT-US02-14341-14 Sequence 14, Appl
35 92 3.4 828 5 US-09-614-150A-36534 Sequence 36534, A
36 92 3.4 1590 7 US-60-495-114-1488 Sequence 1488, Ap
37 90.5 3.4 557 7 US-60-500-337-2394 Sequence 2394, Ap
38 90.5 3.4 580 1 PCT-US03-28227-4205 Sequence 4205, Ap
39 90.5 3.4 580 1 PCT-US03-28227-4206 Sequence 4206, Ap
40 90.5 3.4 580 1 PCT-US03-28227-4207 Sequence 4207, Ap
41 90.5 3.4 580 7 US-60-500-337-2393 Sequence 2393, Ap
42 90.5 3.4 580 7 US-60-500-337-2396 Sequence 2396, Ap
43 90 3.4 665 5 US-09-614-150A-39579 Sequence 39579, A
44 90 3.4 1000 6 US-10-275-762-36 Sequence 36, Appl
45 90 3.4 1026 1 PCT-US03-38193-2539 Sequence 2539, Ap

ALIGNMENTS

RESULT 1
US-09-614-150A-38841
; Sequence 38841, Application US/09614150A
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CLO00728
; CURRENT APPLICATION NUMBER: US/09/614,150A
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38841
; LENGTH: 891
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-614-150A-38841

Query Match 41.4%; Score 1111; DB 5; Length 891;

Best Local Similarity 43.4%; Pred. No. 5.8e-75;

Matches 234; Conservative 75; Mismatches 162; Indels 68; Gaps 14;

QY 3 GMKVEVLNSD-----AVLPSRV---YWIASVIQTAGYRVLLRYEGFENDASHDFWNCILGT 54
Db 218 GMKVEVENTCDSEVLPQPGTPTSFVAVILEIKYKALMSYEGFTD-SHDFWNLN 276
QY 55 VDHPHIGWCAINSKILVPPRTTHAKFTDWKGYLMKRLVGSRTLPVDPHKKMVSMTKPPFR 114
Db 277 AEVHSVQCATRGKPLIPPTTIEHKYKMDKDFLVGLSGARTLPSPFNKINDSLQSRPR 336
QY 115 QGMELEVVDKRSQVSRTRMAVVDVTIGGRLLLYEDGSDDDDFWCHMWSPLIHPVGSRRV 174

```

Db 337 LGNLECVKDRISQVRLATVTKIVGKRLFLRY--FDSDDGFCHEDSPIIHPVGWATTV 394
Qy 175 GHGI-----KMSERRSDMAHPTFRKIYCDVAPYLFK---KYRAVYTEG--GWFEF 220
Db 395 GHNLAAQODYLERMLAGREAMIEVHE-----DATIELFKMFTFDEYSDGKTNSEFVE 448
Qy 221 GMKLEAIDPLNLGNICVATVCKVLLDGYLMICVDG--GPSTDGLDWFCHASSHAIFPATF 279
Db 449 GMKLEAVDPLNLSICPATVMAVLKFGYMMIRIDSYQPDASGSDWFCYHKSPCIFPAGF 508
Qy 280 CQKNDIELTPKGYEAQTFNWNYLEKTSKAAPSLFNMDCPNHGFKVGMKLEAVDLME 339
Db 509 CSVNNISVTPNGYDSRTFTWEGYLRDTGAVAAQHLFHRIIIDHGFEGVMSLECADLMD 568
Qy 340 PRLICVATVRVVRHLLSIHFDGWDSEYDQWDCESPDYVPGWCBLTGQYQLOPP----- 394
Db 569 PRLCVATVARVVRGRLKLVHFDGWDTEYDQWDCESADYVPGVCLVNVHKLGGPPRAV 628
Qy 395 -VAAEPATPLKAKEATKKKKQFGKKRKPPTKTRPLRQSGKKPLEDPOGARKISSE 453
Db 629 QOAPKPAKPKIORKKRPKGAAGK-----TPDNNNTQSVKSRITALKTTTPH----- 676
Qy 454 PVPGELIATVRVKEHLDA-----SPDKASSPELP-----VSVENIKQETDD 495
Db 677 -LPKLSIKLEKPEHNAAFYENNQPEEGDEEDPDADGCGDSTSHISEQSTQSSD 734

```

RESULT 2

```

US-09-614-150A-27069
; Sequence 27069, Application US/09614150A
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/09/614,150A
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27069
; LENGTH: 1243
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-614-150A-27069

```

```

Query Match 41.4%; Score 1111; DB 5; Length 1243;
Best Local Similarity 43.4%; Pred. No. 8.1e-75;
Matches 234; Conservative 75; Mismatches 162; Indels 68; Gaps 14;

Qy 3 GMKVEVLNSD-----AVLPSRV---YWIASVIQTAGVRLVLYEGFENDASHDFWNCILCT 54
Db 570 GMKVEVLNSDSEIVTQPGTPTSFVWATILITTCQQLILRYCGYGEDRRADFWCDVVIADLHPVG 628
55 VDVHPIGWCAINSKILVPPRTIHAFTDWKGYLMKRLVGSRTLPVDFFHIMVSMKYPFR 114

```

```

Db 629 AEVHSVGWCATRGKPLIPPTTIEHKYKWDKDFVGRLSGARTLPSNFYNKINDSLOSRFR 688
Qy 115 QGMLEVVVDKSOVSRRTMAVVDTVIGRLLLYEDGSDDDFWCHMWSPLIHPVGNRRY 174
Db 689 LGNLECVKDRISQVRLATVTKIVGKRLFLRY--FDSDDGFCHEDSPIIHPVGWATTV 746
Qy 175 GHGI-----KMSERRSDMAHPTFRKIYCDVAPYLFK---KYRAVYTEG--GWFEF 220
Db 747 GHNLAAQODYLERMLAGREAMIEVHE-----DATIELFKMFTFDEYSDGKTNSEFVE 800
Qy 221 GMKLEAIDPLNLGNICVATVCKVLLDGYLMICVDG--GPSTDGLDWFCHASSHAIFPATF 279
Db 801 GMKLEAVDPLNLSICPATVMAVLKFGYMMIRIDSYQPDASGSDWFCYHKSPCIFPAGF 860
Qy 280 CQKNDIELTPKGYEAQTFNWNYLEKTSKAAPSLFNMDCPNHGFKVGMKLEAVDLME 339
Db 861 CSVNNISVTPNGYDSRTFTWEGYLRDTGAVAAQHLFHRIIIDHGFEGVMSLECADLMD 920
Qy 340 PRLICVATVRVVRHLLSIHFDGWDSEYDQWDCESPDYVPGWCBLTGQYQLOPP----- 394
Db 921 PRLCVATVARVVRGRLKLVHFDGWDTEYDQWDCESADYVPGVCLVNVHKLGGPPRAV 980
Qy 395 -VAAEPATPLKAKEATKKKKQFGKKRKPPTKTRPLRQSGKKPLEDPOGARKISSE 453
Db 981 QOAPKPAKPKIORKKRPKGAAGK-----TPDNNNTQSVKSRITALKTTTPH----- 1028
Qy 454 PVPGELIATVRVKEHLDA-----SPDKASSPELP-----VSVENIKQETDD 495
Db 1029 -LPKLSIKLEKPEHNAAFYENNQPEEGDEEDPDADGCGDSTSHISEQSTQSSD 1086

```

RESULT 3

```

US-10-719-993-647
; Sequence 647, Application US/10719993
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 647
; LENGTH: 894
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-647

```

```

Query Match 22.8%; Score 611; DB 6; Length 894;
Best Local Similarity 34.1%; Pred. No. 1.1e-37;
Matches 156; Conservative 65; Mismatches 170; Indels 66; Gaps 15;

Qy 3 GMKVEVLNSDAVLSRVVYWIASVIQTAGVRLVLYEGFENDASHDFWNCILGTVDVHPIG 62
Db 77 GMKLEVANKN---NPDTYVWATILITTCQQLILRYCGYGEDRRADFWCDVVIADLHPVG 133
Qy 63 CAINSKILVPPRTIHAFTDWKGYLMKRLVGSRTLPVDFFHIMVSMKYPFR----- 114
Db 134 CTQNNKVLMPDPAIKEKYTDWTEFLRLDTGSRTPANL-----LEGPLRGKGPIDL 186
Qy 115 -QGMLEVVVDKSOVSRRTMAVVDTVIGRLLLY---EDGSDDDFWCHMWSPLIHPVGN 170
Db 187 TVGSLIELQDSQNPQYVIVSVIENVGRLRLRVGLDETSYDQ--WLFYLDVYKRLRVGW 245
Qy 171 SRRVGHGKIMSERSD-----MAHH--PTPRKIYCDVAPY-----LFFKVRAYYTE 214
Db 246 CQE-----NKYRWDPPSEIYPLKMASEWKCULEKSLIDAAKFFPLPMEVFKHADLRSH 298
Qy 215 GGMFEFGKLEAIDPLNLGNICVATVCKVLLDGYLMICVDG--GPSTDGLDWFCHASSHA 273
Db 299 --FTVGMKLETVNMCEFFYIISPASVTKVFNHFFQVTIDDLRPEPSKLSMLC--HADS 355

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2004, 09:05:51 ; Search time 20 Seconds

(without alignments)
2380.175 Million cell updates/sec

Title: US-10-031-915-36

Perfect score: 2883

Sequence: 1 MKGMKVEVLNSDAVLPSRVY.....KASSPELPVSVENIKQETDD 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	494.5	18.4	619	2 T23964	hypothetical prote
2	471	17.6	711	2 T12525	hypothetical prote
3	402	15.0	1477	2 T13797	tumor suppressor pr
4	384.5	14.3	390	2 T14794	hypothetical prote
5	116.5	4.3	200	2 F64217	ribosomal protein
6	104	3.9	6669	2 S55024	nebulin, skeletal
7	99.5	3.7	1314	2 T09481	mating type silenc
8	98	3.7	871	2 G86586	DNA topoisomerase
9	98	3.7	871	2 D72038	DNA topoisomerase
10	97.5	3.6	739	2 S18642	polynucleotide ade
11	97.5	3.6	1379	2 T45119	FIM protein [impor
12	97	3.6	689	2 S17875	polynucleotide ade
13	96.5	3.6	740	2 S17925	polynucleotide ade
14	95.5	3.6	721	2 A33319	microtubule-associ
15	95	3.5	322	2 B43427	neurofilament prot
16	94	3.5	756	2 T49475	related to tol pro
17	94	3.5	1644	2 F91286	hypothetical prote
18	94	3.5	1644	2 B86128	hypothetical prote
19	93	3.5	701	2 F84912	hypothetical prote
20	93	3.5	1113	2 T47381	hypothetical prote
21	92.5	3.4	463	2 T39004	probable histone a
22	92.5	3.4	526	1 T45058	phosphoprotein pho
23	92	3.4	699	2 T09485	Cys-rich protein R
24	92	3.4	815	2 JG0197	myosin-light-chain
25	92	3.4	869	2 H86440	unknown protein [i
26	91.5	3.4	246	2 S26826	histone H1 - maize
27	91.5	3.4	1285	2 H85041	hypothetical prote
28	91.5	3.4	1493	2 F70435	glutamate synthase
29	91	3.4	401	2 S54437	hypothetical prote

30	91	3.4	486	2 S07633	hematopoietic line
31	91	3.4	574	2 T27100	hypothetical prote
32	90.5	3.4	803	2 T18738	hypothetical prote
33	90.5	3.4	1314	2 H83327	protein F180A4.25
34	90.5	3.4	1331	2 T49813	related to gastrin
35	90.5	3.4	2150	1 S27802	zinc finger protei
36	90.5	3.4	2150	2 T19450	hypothetical prote
37	90	3.4	207	2 T07035	histone H1, stress
38	90	3.4	1020	1 QF0703	neurofilament trip
39	89.5	3.3	510	2 JS0724	cytochrome P450 AL
40	89.5	3.3	1162	2 E84431	probable Na+/H+ an
41	89	3.3	806	2 A43427	neurofilament trip
42	89	3.3	1012	2 JC5925	membrane klotho pr
43	89	3.3	3924	2 S37431	ankyrin 2, neurona
44	88.5	3.3	929	2 T35683	ftsk homolog - Str
45	88	3.3	202	2 S65059	histone H1, drough

ALIGNMENTS

RESULT 1

T23964 Hypothetical protein R06C7.7 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C/Accession: T23964

R/Gardner, A.

submitted to the EMBL Data Library, April 1996

A/Reference number: Z19825

A/Accession: T23964

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-619 <WIL>

A/Cross-references: EMBL:Z71266; PIDN:CAA95838.1; GSPDB:GN00019; CESP:R06C7.7

A/Experimental source: clone R06C7

C/Genetics:

A/Gene: CESP:R06C7.7

A/Map position: 1

A/Introns: 121/3; 182/3; 238/2; 270/3; 547/3; 594/2

Query Match					18.4%; Score 494.5; DB 2; Length 619;
Best Local Similarity					29.7%; Pred. No. 2.1e-33;
Matches 132; Conservative 81; Mismatches 149; Indels 83; Gaps 18;					
QY	10	NSDAVLPSPVYVIAVIQTAGYRVLLRYEGENDASHDFWNLGTVDPVHPICWCAINSK- 68			
DB	193	NCDSI---QVRFAIEKVCGYRVLAQFIG---ADTKFWLNI-SDDMFGLANAAMSDPN 245			
QY	69	---ILVPPRTTHAKF-TDWKGYLMK---RLVGSRTLPVDFHIKWVESMKYPPFQGMRL 119			
DB	246	MDKIYVAPPLAINEEYQNDWYNNVNCIDGIEIVGQTSLSPKFDEGKALLSKHFKVGQRL 305			
QY	120	EVVDKSVSRTRMAVVDVTVIGRLR--LLYED-----GSDDDD-----FWCHMWSP 163			
DB	306	ELLNYSNSTEIRVARIQICRRMNVSTIKDFPESLPDADDQRFSSGSQGYWIDEGSF 365			
QY	164	LHHPVGM-----SRRVGHGIKNSERSDMAHPTFRKIYCDV-PYLF 205			
DB	366	FIFPGFRAVNGYQLNAKEYIEHTNKIAQIKGENPRYSDDVTFDQLAKDPIDPMIW 425			
QY	206	KKRVAVYTEGGFESGMKLEAIDPL--NLGNCIVATV---CKVLDDGYLMICVDGPGSTD 260			
DB	426	RKVKV-----GQFELIDPLAQQFNHLVASILKFKCT--EGYLVGMGDPDALE 473			
QY	261	GLDWFCYHASSHAIFPATFCQKNDIELTPKGYEAQTENWENYLEKTSKAAPSELFNMD 320			
DB	474	--DSFPFIHNTFMPVGVAEKYNLELVPDFEFG-TFRWDEYLEKESAETLPLDLFK-P 529			
QY	321	CPNH-----GFKVGMKLEAVDLMPEPLICVATVKRVVHRLLSIHFDGMDSEYDQ 369			
DB	530	MPSQERLDKFKVILLSKVGRLRLAADNCENQFICPATVKSVMHGLINNVNFDGWDSEFDE 589			

QY 370 WVDCESPDIYPVWGCELTGYQLQPP 394
Db 590 LYVDVSHDILPIGWCEAHSVVLQPP 614

RESULT 2
T12525
hypothetical protein DKFZp434L243.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 23-Jul-1999
C:Accession: T12525
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z17524
A:Accession: T12525
A:Molecule type: mRNA
A:Status: preliminary
A:Residues: 1-711 <WAM>
A:Cross-references: EMBL:AL080140
A:Experimental source: adult testis; clone DKFZp434L243
C:Genetics:
A:Note: DKFZp434L243.1

Query Match 17.6%; Score 471; DB 2; Length 711;
Best Local Similarity 30.2%; Pred. No. 2.4e-31;
Matches 133; Conservative 79; Mismatches 186; Indels 42; Gaps 14;

QY 66 NSKILVPPRTIHAFTDKWGYLMKRLVGSRTLPVDFHIKVMKYP---FROGMELEV 122
Db 1 NKTLEAPGIRDKVSDWDEFLRGLTIGACSPVPL-LEGLRGRNPLDLIAPGRLEQ 59

QY 123 DKSQVSRTRMAVDTVIGRLRLIYEDGSDDDP--WCHMWSPLIHPVGSRRVGHGIXM 180
Db 60 AFQSLSTWIVTVVEN:GGRUKRYEGLESNDYEHWLYLDPLFHHVWAAQQGYELQP 119

QY 181 SERRSDMAHPTFRKIY-----DAPV-YLF--KKVAVYTEGGWFEEGMKLEAIDPL 230
Db 120 PSAIRHLKNEAEWQELAKVKEEELPSYLFKDKQVIGHT---FSVNMKLEAVDPW 175

QY 231 NLGNICVATVKVLLDGYLMICVUG-GPSTDGLDFWFCYHASSHAIPATFCQNDIELTP 289
Db 176 SPFGISPATVVKVDFEYFLVEMDDLPRFNHARRSFVCHADSPGIFPVQMSLKNGLHIS 235

QY 290 PKGYEACOTFNWENYLETKSKAAPSRLFNMDCPNHGPKVGMKLEAVDLMPEPRLICVATVK 349
Db 236 PPGYPSQDFDADVLKCGAEEAQRCPPLISHEFEKENKLEAVNPILPEEVCVATIT 295

QY 350 RVVHRLLSIHDPGWDSEYDQW-VDCESPDIYPVWGCELTGYQLQPPVAAEPATPLKAKEA 408
Db 296 AVRGSYLWLQLESGKPIPECIVSVESMDIFPLGCETNGHPL-----STPRRAR-V 346

QY 409 TKKKKKQFGKKRIPPTKTRPLQSGKKPLEDDPGARKISSEPVPGIIVRVKKEH 468
Db 347 YKQRIAVVQEKQVPSRTI--VHEGLN-----QELNSTSVMLNGKYCCPKYFNH 397

QY 469 LDVASP--DKASSPELPSV 486
Db 398 RCFSGPYLNGRIAEPLQCV 417

RESULT 3
T13797
tumor suppressor protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13797
R:Wismar, J.; Loeffler, T.; Habtemichael, N.; Vef, O.; Geissen, M.; Zirwes, R.; Altmeyer
Mech. Dev. 53, 141-154, 1995
A:Title: The Drosophila melanogaster tumor suppressor gene lethal(3)malignant brain tumor
A:Reference number: Z17760; MUID:96100389; PMID:8555106
A:Accession: T13797
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA

A:Residues: 1-1477 <WIS>
A:Cross-references: EMBL:X80839; NID:G530289; PIDN:CAA56811.1; PID:G3421009
C:Genetics:
A:Cross-references: FlyBase:FBgn0002441
A:Introns: 210/3; 746/1

Query Match 15.0%; Score 402; DB 2; Length 1477;
Best Local Similarity 21.9%; Pred. No. 4e-25;
Matches 117; Conservative 83; Mismatches 159; Indels 176; Gaps 16;

QY 3 GMKVEVLSDAVLPSRVVWIASVIQTAGYRVLLVEGFENDASHDFWCNLGTVDVHPHIGW 62
Db 853 GMKLEAIDPE---NCSLFCVCSIVVEGRYLUKGSFDG---SSMYDFWVWADSQDIFPPGW 907

QY 63 CAINKSLVPPRTIHAFTDKWGYLMKRLVGSRTLP--VDFHIKVMKYSM--KYFPQGM 118
Db 908 CDETARVLOAFKDYNSERSFSRVLVK--TGKAAAPRALFGLHANNQQQMDVRENGFAVGMH 965

QY 119 LEVDVKSQVSRTRMAVDTVIGRLRLIYEDGSDDDDFWCHMWSPLIHPVGSRRVGHG 178
Db 966 LEAEDLNDTGKICVATVTDILDERIRVHFQGWDDCYDLWVHITSFYTHPCGW----- 1017

QY 179 KMSERRSDMAHPTFRKIYCDAPVYLFKKRAVYTEGGWFEEGMKLEAIDPLNLGNICVA 238
Db 1018 -----HEGRQ----- 1022

QY 239 TVCKVLLDGYLMICVDGSPSTGLDFWFCYHASSHAIPATFCQNDIELTPPKGYEATF 298
Db 1023 -----QLIVPPDYQKSAF 1035

QY 299 NWENYLETKSKAAPSRLFNMDCPNHGPKVGMKLEAVDLMPEPRLICVATVKRVHRLLSI 358
Db 1036 INDDYIISVGGVNAKSELFTPRQME-YQERMKLEVDVDQNPCLIRPATVTVTRKGYRVL 1094

QY 359 HPDGDWSDYDQWVDCESPDIYPVWGCELTGYQLQ-PPVAAEPATPLKA----- 405
Db 1095 HLDGWPTFYVWLEDDSDPLHPIGWCEATSHLETPPGYLOPKSVMPDCDEGCRGFGNAK 1154

QY 406 -----KEATKKKKQFGKKR-KRIPPTKTRP--LRQG-----SKKPLED 442
Db 1155 RPNLNVHRLRECCYPANRWQWASKTVKPPRVAPENIRRWAKTKKACSEAKQAIKD 1214

QY 443 DPQGARKISSEPVPGIIVRVKSEHLDVASPDKASSP--ELPVSVENIKQETDD 495
Db 1215 SQQ-----EIVYPKVAE-----VQAKRKTSPCEEKWKVKQKQMKED 1253

RESULT 4
T14794
hypothetical protein DKFZp586P1522.1 - human
C:Species: Homo sapiens (man)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14794
R:Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, August 1999
A:Reference number: Z18180
A:Accession: T14794
A:Molecule type: mRNA
A:Status: preliminary
A:Residues: 1-390 <KOB>
A:Cross-references: EMBL:AL110279
A:Experimental source: adult uterus; clone DKFZp586P1522
C:Genetics:
A:Note: DKFZp586P1522.1

Query Match 14.3%; Score 384.5; DB 2; Length 390;
Best Local Similarity 36.2%; Pred. No. 2e-24;
Matches 94; Conservative 32; Mismatches 95; Indels 39; Gaps 9;

QY 222 MKLEAIDPLNLGNICVATVKVLLDGYLMICVDGSPSTGLDFWFCYHASSHAIPATFCQ 281
Db 1 MKLEAVDRMNPISLVCVASVTDV-VDSRELHFDNWDOT--YDYWC-DFSSPYIHPVGMQ 56

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2004, 09:05:51 ; Search time 39 Seconds
(without alignments)

3275.283 Million cell updates/sec

Title: US-10-031-915-36

Perfect score: 2683

Sequence: 1 MKGMKVEVLNSDAVLPSRVY.....KASSPELPVSVENIKOETDD 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp Vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2426	90.4	703	11 Q8BHD5	Q8bhd5 mus musculus
2	1418	52.9	464	11 Q8VE12	Q8ve12 mus musculus
3	1398	52.1	411	11 Q8BZY2	Q8bzy2 mus musculus
4	1395	52.0	410	4 Q9XU1	Q9xui1 homo sapien
5	1111	41.4	1243	5 Q9VK33	Q9vk33 drosophila
6	1079	40.2	359	11 Q8C1B4	Q8c1b4 mus musculus
7	620.5	23.1	866	4 Q96C73	Q96c73 homo sapien
8	620.5	23.1	866	4 Q9UHJ3	Q9uhj3 homo sapien
9	611	22.8	904	4 Q9HCF5	Q9hcf5 homo sapien
10	604	22.5	863	11 Q9JND1	Q9jnd1 mus musculus
11	599	22.3	863	11 Q9JND2	Q9jnd2 rattus norv
12	598	22.3	591	11 Q8CFS1	Q8cfs1 mus musculus
13	525	19.6	781	4 Q96JW7	Q96jw7 homo sapien
14	518.5	19.3	883	11 Q9SLB7	Q9slb7 mus musculus
15	510	19.0	772	4 Q8IU7	Q8iuv7 homo sapien
16	508	18.9	612	5 Q21769	Q21769 caenorhabdi

17	499	18.6	528	6 Q95LL9	Q95ll9 macaca fasc
18	495.5	18.5	534	4 Q8IXS3	Q8ixs3 homo sapien
19	495.5	18.5	623	4 Q8NA19	Q8na19 homo sapien
20	471	17.6	711	4 Q9Y4O9	Q9y4o9 homo sapien
21	402	15.0	1477	5 Q9VBS2	Q9vbs2 drosophila
22	402	15.0	1477	5 Q76931	Q76931 drosophila
23	357.5	13.3	591	4 Q9UKM6	Q9ukm6 homo sapien
24	357.5	13.3	660	4 Q96GD3	Q96gd3 homo sapien
25	353.5	13.2	577	4 Q9UKM5	Q9ukm5 homo sapien
26	353.5	13.2	599	4 Q8WU48	Q8wu48 homo sapien
27	351	13.1	700	4 Q9UQRO	Q9uqro homo sapien
28	349	13.0	664	11 Q9JME0	Q9jme0 mus musculu
29	349	13.0	706	11 Q8K214	Q8k214 mus musculu
30	340	12.7	845	11 Q8BYC8	Q8byc8 mus musculu
31	332	12.4	544	5 Q9N3Q7	Q9n3q7 caenorhabdi
32	330	12.3	722	11 Q99MM4	Q99mm4 mus musculu
33	327	12.2	877	5 Q9VHA0	Q9vha0 drosophila
34	326.5	12.2	250	13 Q9W6R7	Q9w6r7 fugu rubrip
35	323.5	12.1	212	4 Q9NTW3	Q9ntw3 homo sapien
36	289	10.0	114	11 Q8BKW1	Q8bkw1 mus musculu
37	255.5	9.5	173	11 Q8BHZ8	Q8bhz8 mus musculu
38	252.5	9.4	183	11 Q8BG70	Q8bg70 mus musculu
39	226.5	8.4	261	4 Q8IWD0	Q8iwd0 homo sapien
40	221	8.2	405	11 Q8BMN8	Q8bmna8 mus musculu
41	206	7.7	81	4 Q75996	Q75996 homo sapien
42	197	7.3	119	4 Q86LH0	Q86lh0 homo sapien
43	158	5.9	178	4 Q8WX46	Q8wx46 homo sapien
44	130	4.8	302	11 Q8C0G4	Q8c0g4 mus musculu
45	121	4.5	5146	6 Q8SPM4	Q8spm4 bos taurus

ALIGNMENTS

RESULT 1

Q8BHD5

ID Q8BHD5 PRELIMINARY; PRT; 703 AA.

AC Q8BHD5; 01-MAR-2003 (TRENBLrel. 23, Created)

DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

DE H-L.

OS Mus musculus (Mouse)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]_TaxID=10090;

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Bone, and Skin;

RX MEDLINE=22354583; PubMed=12466851;

RA The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs."

RL Nature 420:563-573 (2002).

DR EMBL; AK029115; BAC26305.1; --

DR EMBL; AK036510; BAC29456.1; --

SQ SEQUENCE 703 AA; 78970 MW; 259B5DF5BFCEAFF9 CRC64;

Query Match

Best Local Similarity 90.4%; Score 2426; DB 11; Length 703;

Matches 443; Conservative 23; Mismatches 27; Indels 2; Gaps 1;

QY 1 MKGMKVEVLNSDAVLPSRVYVIAVIQAGYRVLLRYEGFENDASHDFWCLNLTVDVHPI 60

Db 211 MKGMKVEVLNSDAVLPSRVYVIAVIQAGYRVLLRYEGFENDASHDFWCLNLTVDVHPI 270

QY 61 GWCAINSILVPPRTTHAKFTDQWGYLMKRLVGSRTLPVDFPHIKQVSMKYPFQGMRL 120

Db 271 GWCAINSILVPPRTTHAKFTDQWGYLMKRLVGSRTLPVDFPHIKQVSMKYPFQGMRL 330

QY 121 VVDKSQVSRTRMAVVDVTVGGRLLLYEDGSDDDPWCHMWSPLTHPVGVSRRVGHG1KM 180

```
Db 331 VDKTQVSRTRMAVVDVTIGRLRLLEDGSDDDDFWCHMWSPLIHPVGSRRVGHGKM 390
Qy 181 SERSDMAHPTFRKIYCDVAVPYLFKKVRAVYTEGGWFEGMKLEALDPLNLGNICVATV 240
Db 391 SDRCDMSHPTRFKIYCDVAVPYLFKKVRAVYTEGGWFEGMKLEALDPLNLGSI CVAT 450
Qy 241 CKVLLDGYLMTVDGSGSTGLDWFCHYASHAIFPAFFCQKNDIELTPPKGYEAQTFFW 300
Db 451 CKVLLDGYLMTVDGSGSTGSDWFCVHASSHAIFPAFFCQKNDIELTPPKGYETQFFW 510
Qy 301 ENYLEKTSKAAPSRLENMDCPNHGFVKVGMKLEAVDLMEPRLICVATVKRVHRLLSIHF 360
Db 511 EYLEKTSKAAPARLENMDCPNHGFVKVGMKLEAVDLMEPRLICVATVKRVHRLLSIHF 570
Qy 361 DGWSEYDQWDCESPDIYPVGCWELTGYQLQPPVAAEPATPLKAKEATKKKKQFGK 420
Db 571 DGWSEYDQWDCESPDIYPVGCWELTGYQLQPPVAAEPATPLKAKEATKKKKQFGK 630
Qy 421 KRIPPTKTRPLROGSKPILDDPOGARKISSPVPGEIIVAVRKEHLDVASPKASSP 480
Db 631 KRIPPTKTRPLROGSKPILDDPOGARKISSPVPGEIIVAVRKEHLDVASPKASSP 688
Qy 481 ELPSVENIKQETDD 495
Db 689 QLPLPIESIKQERN 703

RESULT 2
Q8BZY2 PRELIMINARY; PRT; 464 AA.
ID Q8BZY2 AC Q8BZY2
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Similar to hypothetical protein FL20055.
GN AA08199.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC020018; AA020018.1; -.
DR MGD; MGI:2143977; AA08199.
DR InterPro; IPR004092; Mdt.
DR Pfam; PF02820; mdt; 4.
DR SMART; SM00561; Mdt; 4.
KW Hypothetical protein.
SQ SEQUENCE 464 AA; 52721 MW; 31E5C041592BE874 CRC64;

Query Match 52.9%; Score 1418; DB 11; Length 464;
Best Local Similarity 56.4%; Pred. No. 11e-126;
Matches 264; Conservative 67; Mismatches 121; Indels 16; Gaps 4;

Qy 2 KGMKVEVLSDAVPSRVYASVIQTAGYRVLLRYEGFENDASHDFWNLGTVDVHPIG 61
Db 10 ENRVIEVPNTDCLPTKVFVIAGIILAGYNALLRYEGFENDSSLDFFWNCIGSDIHPVG 69
Qy 62 WCAINSILVPPRTIHAFTDMKGYLMKRLVGSRTLPLVDPHIKMVESMKYPFRQGRLEV 121
Db 70 WCAASGKPLVPRTVQHKYTNKAFVLKRLTGAKTLPPDFSQKVSQVSEMQYFFPKCMRVEV 129
Qy 122 VDKSQRTRMAVVDVTIGRLRLLEDG-DSDDDFWCHMWSPLIHPVGSRRVGHGKM 180
Db 130 VDKRHLCTRVAVVEVIGRLRLVYEESEDRDFFWCHMWSPLIHPVGSRRVGHGKM 189
Qy 181 SERSDMAHPTFRKIYCDVAVPYLFKKVRAVYTEGGWFEGMKLEALDPLNLGNICVATV 240
Db 190 SDITKQDGH-----FDTPLPAKVEVDOSGEWFGWFKLEALDPLNLSTICVATI 242
Qy 241 CKVLLDGYLMTVDGSGSTGLDWFCHYASHAIFPAFFCQKNDIELTPPKGYEAQTFFW 300
Db 243 RKVLADGFLMIGDGEAAGDSWFCVHASSHAIFPAFFCQKNDIELTPPKGYETQFFW 302
Qy 301 ENYLEKTSKAAPSRLENMDCPNHGFVKVGMKLEAVDLMEPRLICVATVKRVHRLLSIHF 360
Db 303 FDLRTGSIAPVVKLFNKKVDPVNHGFRVGMKLEAVDLMEPRLICVATVTIRHRLRIHF 362
Qy 361 DGWSEYDQWDCESPDIYPVGCWELTGYQLQPPVAAEPATPLKAKEATKKKKQFGK 420
Db 363 DGWSEYDQWDCESPDIYPVGCWELTGYQLQPPVAAEPATPLKAKEATKKKKQFGK 463
```

```
Db 243 RKVLADGFLMIGDGEAAGDSWFCVHASSHAIFPAFFCQKNDIELTPPKGYETQFFW 302
Qy 301 ENYLEKTSKAAPSRLENMDCPNHGFVKVGMKLEAVDLMEPRLICVATVKRVHRLLSIHF 360
Db 303 FDLRTGSIAPVVKLFNKKVDPVNHGFRVGMKLEAVDLMEPRLICVATVTIRHRLRIHF 362
Qy 361 DGWSEYDQWDCESPDIYPVGCWELTGYQLQPPVAAEPATPLKAKEATKKKKQFGK 419
Db 363 DGWSEYDQWDCESPDIYPVGCWELTGYQLQPPVAAEPATPLKAKEATKKKKQFGK 422
Qy 420 KRIPPTKTRPLROGSKPILDDPOGARKISSPVPGEIIVAVRKEHLDVASPKASSP 467
Db 423 HKMTTSQL-----KEELLDGEDYSFLHGASQESNGSATVYIKOE 463

RESULT 3
Q8BZY2 PRELIMINARY; PRT; 411 AA.
ID Q8BZY2 AC Q8BZY2
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Weakly similar to H-L (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC020018; AA020018.1; -.
DR MGD; MGI:2143977; AA08199.
DR InterPro; IPR004092; Mdt.
DR Pfam; PF02820; mdt; 4.
DR SMART; SM00561; Mdt; 4.
KW Hypothetical protein.
SQ SEQUENCE 411 AA; 47002 MW; F0606450B3C926FC CRC64;

Query Match 52.1%; Score 1398; DB 11; Length 411;
Best Local Similarity 52.9%; Pred. No. 7.5e-125;
Matches 249; Conservative 56; Mismatches 83; Indels 8; Gaps 2;

Qy 2 KGMKVEVLSDAVPSRVYASVIQTAGYRVLLRYEGFENDASHDFWNLGTVDVHPIG 61
Db 10 ENRVIEVPNTDCLPTKVFVIAGIILAGYNALLRYEGFENDSSLDFFWNCIGSDIHPVG 69
Qy 62 WCAINSILVPPRTIHAFTDMKGYLMKRLVGSRTLPLVDPHIKMVESMKYPFRQGRLEV 121
Db 70 WCAASGKPLVPRTVQHKYTNKAFVLKRLTGAKTLPPDFSQKVSQVSEMQYFFPKCMRVEV 129
Qy 122 VDKSQRTRMAVVDVTIGRLRLLEDG-DSDDDFWCHMWSPLIHPVGSRRVGHGKM 180
Db 130 VDKRHLCTRVAVVEVIGRLRLVYEESEDRDFFWCHMWSPLIHPVGSRRVGHGKM 189
Qy 181 SERSDMAHPTFRKIYCDVAVPYLFKKVRAVYTEGGWFEGMKLEALDPLNLGNICVATV 240
Db 190 SDITKQDGH-----FDTPLPAKVEVDOSGEWFGWFKLEALDPLNLSTICVATI 242
Qy 241 CKVLLDGYLMTVDGSGSTGLDWFCHYASHAIFPAFFCQKNDIELTPPKGYEAQTFFW 300
Db 243 RKVLADGFLMIGDGEAAGDSWFCVHASSHAIFPAFFCQKNDIELTPPKGYETQFFW 302
Qy 301 ENYLEKTSKAAPSRLENMDCPNHGFVKVGMKLEAVDLMEPRLICVATVKRVHRLLSIHF 360
Db 303 FDLRTGSIAPVVKLFNKKVDPVNHGFRVGMKLEAVDLMEPRLICVATVTIRHRLRIHF 362
Qy 361 DGWSEYDQWDCESPDIYPVGCWELTGYQLQPPVAAEPATPLKAKEATKKKKQFGK 396
Db 363 DGWSEYDQWDCESPDIYPVGCWELTGYQLQPPVAAEPATPLKAKEATKKKKQFGK 398
```

RESULT 4

Q9XU1 PRELIMINARY; PRT; 410 AA.
 ID Q9XU1
 AC Q9XU1
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein FLJ20055.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Ota T., Suzuki Y., Ohyashi M., Nishi T., Shibahara T.,
 RA Tanaka T., Nakamura Y., Isegai T., Sugano S.;
 RT "NEDO human cDNA sequencing project";
 RL Submitted (FE3-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK000082; BAA0919.1; -;
 DR InterPro; IPR004092; Mbt.
 DR Pfam; PF02820; mbt; 4.
 DR SMART; SM00561; Mbt; 4.
 DR KW Hypothetical protein.
 SQ SEQUENCE 410 AA; 46717 MW; 659C09639AC02CF3 CRC64;

Query Match 52.0%; Score 1395; DB 4; Length 410;
 Best Local Similarity 62.9%; Pred. No. 1.4e-124;
 Matches 249; Conservative 55; Mismatches 84; Indels 8; Gaps 2;
 QY 2 KGMKEVLNSDAVPSRVYVIASVIOTAGYVLLRYEGFENDASHDFWNCNLGTVDPVPIG 61
 DB 10 ENVEVEVNTDCSLPTKVFVIAIGLVKLAGTALLRYEGFENDSGLDGFWNCICGSDLHPVG 69
 QY 62 WCAINSILVPPRTIHAFTDWMKLMKRLVSGRTLPVDFPHIKVSMKYPFRQGNRLEV 121
 DB 70 WCAASGKPLVPPRTIQHYKNWKAFLVKRLTGAKTLPPDFSQVSSMQYFPKQPRVEV 129
 QY 122 VDKSQVSRMTAVVDTVIGRLRLLYEDG-DSDDDFCHWNSPLIHPVGHRSRVHGIGM 180
 DB 130 VDKHLCRTIRVAVVESVIGRLRLVVEESDRDFFWCHMSPLIHHIGWSRISGHRFR 189
 QY 181 SERASDMAHPTFRKIYCDAPVPLFKKRVAVYTEGGWFEGMKLEALDPLNLGNCVATV 240
 DB 190 SDITFKQDGH-----FDTPPHLFAKVEVDQSGEWFEGMKLEALDPLNLSTICVAT 242
 QY 241 CKVLIDGLMICVDGPGSTGLDWFCHVASHAIFPATFCQKNDIELTPPKGYEATFNW 300
 DB 243 RKVLADGLMIGIDGSEADGSDMFCVHATSPSIFPVGCEINNIETPPRGYTKLPFKW 302
 QY 301 ENYLEKTSKAAPSRPLFNMDCPNHGFVKMKLEAVDLMEPRLLICVATVKRVHRLLSIF 360
 DB 303 FDYLRETSAAAPKLFNKQVPHNGFRVGMKLEAVDLMEPRLLICVATVTRIHLRLIIF 362
 QY 361 DGDWSEVDQWDCESPIYVGVCELTGYQLQPPVA 396
 DB 363 DGMEEYDQWDCESPLYPVGVGCOLTGYQLQPPAS 398

RESULT 5

Q9VK33 PRELIMINARY; PRT; 1243 AA.
 ID Q9VK33
 AC Q9VK33
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE CG16975 protein.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscophora;

OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Fiankoc C., Baldwin D.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari L., Beasley E.M.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis J.M., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Hartis N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Stradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS. A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 DR EMBL; AE003639; AAF53249.1; -;
 DR EMBL; AE003639; AAF53250.1; -;
 DR FlyBase; FBgn0032475; CG16975.
 DR InterPro; IPR004092; Mbt.
 DR InterPro; IPR001660; SAM.
 DR Pfam; PF02820; mbt; 4.
 DR Pfam; PF00536; SAM; 1.
 DR SMART; SM00561; Mbt; 4.
 DR SMART; SM00454; SAM; 1.
 DR PROSITE; PS0105; SAM DOMAIN; 1.
 DR KW Alternative splicing; Hypothetical protein.
 FT VARSPLOC 1 352 MISSING (IN SHORT ISOFORM).
 SQ SEQUENCE 1243 AA; 136036 MW; 35FC45F6298F5BAA CRC64;

Query Match 41.4%; Score 1111; DB 5; Length 1243;
 Best Local Similarity 43.4%; Pred. No. 9.7e-97;
 Matches 234; Conservative 75; Mismatches 162; Indels 68; Gaps 14;
 QY 3 GMKVEVLNSD-----AVLPSRV---YVIASVIOTAGYVLLRYEGFENDASHDFWNCNLGT 54
 DB 570 GMKVEVENTDCDSIEVTPQGTPTSPFWAILIKGYKALMSYEGEDTD-SHFVWNLGN 628
 QY 55 VDVHPICGCAINSKILVPPRTIHAFTDWMKLMKRLVSGRTLPVDFPHIKVSMKYPFR 114
 DB 629 AEVHVGWCAITRGKPLIPPTIEHKYKWKDFLVGLSGARTLPSNPFYNKINDSLQSRFR 688
 QY 115 QGMRELVVDSQVSRTEMAVVDVTIGRLELLVEDGSDDDDFWCHMSPLIHPVGNRRV 174

```
Db 689 LGLNECVDRISQVRLATVTKVGRFLRY--FSDDDGFMCHEDESPIIHFVGMATTV 746
Qy 175 GHGI-----KMSERRSDMAHPTFRKIYCDVPLFK--KVRVYTEG--GWFER 220
Db 747 GHNLAAQDYLRLMLAGREAVIHEH-----DATIELFKMFTFDEYVSDKTSNFSVE 800
Qy 221 GMLKLEADPLNLGNICVATVCKVLLDGYLMICVDG--GPSTDGLDWCFCYHASSHAIPATF 279
Db 801 GMLKLEADPLNLGICPATVNAVLFYGMIRIDSYQFDASGSDWFCYTHEKSPCIFPAGF 860
Qy 280 CQKNDIELTPKGYEATFNWENYLEKTKSAAAPSLFNMDCPNHFCKVGMKLEAVDLME 339
Db 861 CSVNIISVTPNGVDSRTFTWEGYLRDTGAVACQHLFHRIIIDHGFVGNLSLECADLMD 920
Qy 340 PRLICVATVKKVVRHLLSIHFHGDWSEYDQWDCESPDIYPVGMWCLTGYQLQPP----- 394
Db 921 PRLVCVATVAVVGRLLKLVHFDGWTDEYDQWDCESADIYPVGMWCLVNVHKLGGPPRAVH 980
Qy 395 -VAEPATPLKAKATKKK-KKQFGKKRIPPTKTRPLRQSGKKPLEDDPQGARKISSE 453
Db 981 QOAPKPAKPKIQRKAPKGAAGK-----TPDNTQSVKSRITIAKTPH----- 1028
Qy 454 PVPGEIIAVRVKEHLDVA-----SPDKASSPELP-----VSVENIKQSTDD 495
Db 1029 -LPKLSIKLEKPEHNAAFYENNQPEEGDEEDPDADGDGSGTSHISEQSTTQSSD 1086

RESULT 6
Q8C1B4 PRELIMINARY; PRT; 359 AA.
AC Q8C1B4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Weakly similar to H-L (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RA MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK028503; BAC25982.1; -.
FT NON TER
SQ SEQUENCE 359 AA; 41035 MW; CA233E1B6B60CA04 CRC64;

Query Match 40.2%; Score 1079; DB 11; Length 359;
Best Local Similarity 56.2%; Pred. No. 1.9e-94;
Matches 209; Conservative 45; Mismatches 102; Indels 16; Gaps 4;

Qy 98 PVDPHIKVSMKYPFQGMLEVVDKSKQSRTRMAVVDTVIGRLRLLYEDG--DSDDDF 156
Db 1 PPDPSQVSMQYFPKPCRMVEVDKRLCRTRVAVVESVIGRLRLVYEESEDRTDFF 60
Qy 157 WCHMWSPLIHPVGSRVGHGKMSERRSDMAHPTFRKIYCDVPLFKVAVYTEGG 216
Db 61 WCHMWSPLIHHGWSRSGHFRKRSIDITKKQDGH-----FDPHLLFAKVEVDQSGE 113
Qy 217 WFEGMKLEADPLNLGNICVATVCKVLLDGYLMICVDGSPSTGLDWCFCYHASSHAIPF 276
Db 114 WFEGMKLEADPLNLSTICVATIRKVLADGFLMIGDGEAAGSDWFCYHATSPSIFP 173
Qy 277 ATPCQKNDIELTPKGYEATFNWENYLEKTKSAAAPSLFNMDCPNHFCKVGMKLEAVD 336
Db 174 VGFCEINNIELTPRPGYTKLPFKMFDYLRGTGSIAPVFLFNKDVFNHGFVGMKLEAVD 233
```

```
Qy 337 LMEPLICVATVKKVVRHLLSIHFHGDWSEYDQWDCESPDIYPVGMWCLTGYQLQPPVA 396
Db 234 LMEPLICVATVTRIIRHLLRIHFHGDWSEYDQWDCESPDIYPVGMWCLTGYQLQPPAS 293
Qy 397 AEPATPLKAKATKK-KKQFGKKRIPPTKTRPLRQSGKKPLEDDPQGARKISSEPV 455
Db 294 QSSRESQASGSKQKAKSKQYKHKWMTSQL-----KEELLGDGYSFLHGASQOE 346
Qy 456 PGEIIAVRVKEE 467
Db 347 SNGSATVVIKOE 358

RESULT 7
Q96C73 PRELIMINARY; PRT; 866 AA.
AC Q96C73;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to RUI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1];
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014614; AAI14614.1; -.
DR InterPro; IPR004092; MBT.
DR InterPro; IPR001660; SAM.
DR Pfam; PF02820; mbt; 4.
DR Pfam; PF00536; SAM; 1.
DR SMART; SM00561; MBT; 4.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00454; SAM; 1.
SQ SEQUENCE 866 AA; 98141 MW; DCE67BF35C413EB7 CRC64;

Query Match 23.1%; Score 620.5; DB 4; Length 866;
Best Local Similarity 31.9%; Pred. No. 4.3e-50;
Matches 161; Conservative 91; Mismatches 205; Indels 47; Gaps 16;

Qy 3 GMYEV-LNSDAVLPSRVYTIASVIOTAGYRVLLRYEGFENDASHDFWCNLTVDVHPIG 61
Db 53 GMLKLEAVRTD---PETYVATVITTCBQLLLRYDGYGDRADRWCDIRKADLYPIG 108
Qy 62 WCANSKILVPPRIHAKFTDKGYLMKRLVGSRTLPVDPHIKVSMKYP---FROGMR 118
Db 109 WCEQNKKTLEAPGIRDKVSDWDEFRLQTLTGACSPVPL-LEGLRGRNPLDLIAFGSR 167
Qy 119 LEVVDKSKQSRTRMAVVDTVIGRLRLLYEDGSDDDDF--WCHMWSPLIHPVGSRRVGH 176
Db 168 LECQAFQDSLSTWITVTVVENIGRLKLYRTEGLESSDNYEHWLYYLDPLFHHVGAQQGY 227
Qy 177 GIKMSERRSDMAHPTFRKIYCDVPLFKVAVYTEGGFEGMKLEA 226
Db 228 ELQPPSATIRHLKNAEQETILAKVKEEBEPLSYLFKDKQVIGIHT----FSVNMKLEA 283
Qy 227 IDPLNLGNICVATVCKVLLDGYLMICVDG--GPSTDGLDWCFCYHASSHAIPATCQKNDI 285
Db 284 VDPNSPGCISATVVKVYFDEKYLFLVEMDDLRPENHARRSFVCHADSPGIFPVQMSLKNGL 343
Qy 286 ELTPKGYEATFNWENYLEKTKSAAAPSLFNMDCPNHFCKVGMKLEAVDLMEPRLICV 345
Db 344 HISPPGYPGSDFDWADYLYKQCGAAAPQRCFPLPISHEHEFKENMKLEAVNPILPEEVCV 403
Qy 346 ATVKKVVRHLLSIHFHGDWSEYDQW--VDCESPDIYPVGMWCLTGYQLQPPVAEPATPLK 404
Db 404 ATITAVRGSYLVNLQEGSKPIPECIVSVESMDLPLFGWCTETNGHPL-----STPRR 455
Qy 405 AKEATKKKKQFGKKRIPPTKTRPLRQSGKKPLEDDPQGARKISSEPVGPBIIAVRV 464
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 4, 2004, 16:20:44 ; Search time 883 Seconds
(without alignments)
10658.756 Million cell updates/sec

Title: US-10-031-915-90

Perfect score: 2555

Sequence: 1 agacatccactctatgac.....ttgcttgagaaaaa 2555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2550.8	99.8	2781	13	US-10-117-722-622
2	2550.8	99.8	2781	15	US-10-037-270-622
3	2538.8	99.4	3188	13	US-10-094-749-917
4	2514.4	98.4	2663	12	US-10-264-237-419
5	1227.8	48.1	2432	12	US-10-108-260A-1658
6	512.2	20.0	513	13	US-10-027-632-70682
7	512.2	20.0	513	13	US-10-027-632-70683
8	512.2	20.0	513	14	US-10-027-632-70683
9	512.2	20.0	513	14	US-10-027-632-70683
10	471.2	18.4	472	13	US-10-027-632-293974
11	471.2	18.4	472	13	US-10-027-632-302414
12	471.2	18.4	472	14	US-10-027-632-293974
13	471.2	18.4	472	14	US-10-027-632-302414
14	438	17.1	492	11	US-09-918-995-22234
15	413.8	16.2	3590	9	US-09-789-919-55

16	411.4	16.1	458	11	US-09-918-995-26775	Sequence 26775, A
17	386	15.1	445	11	US-09-918-995-26098	Sequence 26098, A
c 18	324.4	12.7	476	10	US-09-867-701-701	Sequence 701, App
19	236.8	9.3	581	13	US-10-029-386-142	Sequence 142, App
20	236.8	9.3	581	13	US-10-027-632-103678	Sequence 103678, App
c 21	236.8	9.3	2445	13	US-10-027-632-111883	Sequence 111883, App
22	236.8	9.3	2445	14	US-10-027-632-103678	Sequence 103678, App
c 23	236.8	9.3	2445	14	US-10-027-632-111883	Sequence 111883, App
24	235.4	9.2	543	13	US-10-029-386-4056	Sequence 4056, App
25	235.4	9.2	558	9	US-09-864-761-9342	Sequence 9342, App
26	235	9.2	238	13	US-10-029-386-13849	Sequence 13849, A
27	232	9.1	232	9	US-09-864-761-25876	Sequence 25876, A
28	232	9.1	232	13	US-10-029-386-17758	Sequence 17758, A
29	135.8	5.3	165	9	US-09-777-564-1597	Sequence 1597, App
30	135.8	5.3	165	15	US-10-015-219-1597	Sequence 1597, App
c 31	121	4.7	283	12	US-10-242-535A-11074	Sequence 11074, A
32	121	4.7	555	9	US-09-864-761-9308	Sequence 9308, App
33	120.8	4.7	557	9	US-09-864-761-9205	Sequence 9205, App
c 34	117.8	4.6	2635	9	US-09-789-404-1	Sequence 1, Appli
35	117	4.6	159	9	US-09-864-761-25857	Sequence 25857, A
36	106.6	4.2	413	11	US-09-918-995-35932	Sequence 35932, A
c 37	104	4.1	138	12	US-10-242-535A-6856	Sequence 6856, App
38	98	3.8	2099	12	US-10-104-047-1542	Sequence 1542, App
39	98	3.8	2845	12	US-10-104-047-292	Sequence 292, App
40	94.6	3.7	174	9	US-09-789-919-8	Sequence 8, Appli
41	92.2	3.6	237	12	US-10-305-720-588	Sequence 588, App
42	91.8	3.6	586	9	US-09-864-761-9319	Sequence 9319, App
43	91.8	3.6	694	13	US-10-027-632-137841	Sequence 137841, App
44	91.8	3.6	694	13	US-10-027-632-137842	Sequence 137842, App
45	91.8	3.6	694	14	US-10-027-632-137841	Sequence 137841, App

ALIGNMENTS

RESULT 1
US-10-117-722-622
Sequence 622, Application US/10117722
Publication No. US20030219744A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Dmanac, Radoje T.
TITLE OF INVENTION: No. US20030219744A1elel Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 784CIP2BCIP
CURRENT APPLICATION NUMBER: US/10/117,722
CURRENT FILING DATE: 2002-04-04
PRIOR APPLICATION NUMBER: 09/620,312
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 05/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1104
SOFTWARE: pt_FL_genes Version 1.0
SEQ ID NO 622
LENGTH: 2781
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (263)..(1750)
US-10-117-722-622

Query Match 99.8%; Score 2550.8; DB 13; Length 2781;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2552; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AGACATCCACTCTATGACCTGGAGGATGTGATGAAGGATGAAGGTGGAGTGC 60

Db 228 AGCAGTCCCACTTATGACCAAGTGGAGGATGTGTAAGAGGATGAAGGTGGAGGTGC 287
 Qy 61 TCAACAGTGTGTCTCTCCAGCGGGTGTACTGGATCGCTCTGTCTATCCAGACAG 120
 Db 288 TCAACAGTGTGTCTCTCCAGCGGGTGTACTGGATCGCTCTGTCTATCCAGACAG 347
 Qy 121 CAGGATATCGGGTGTCTCTGGTATGAAGGCTTTGAAAATGACGCCACGCTTCT 180
 Db 348 CAGGATATCGGGTGTCTCTGGTATGAAGGCTTTGAAAATGACGCCACGCTTCT 407
 Qy 181 GGTGAACCTTGGGAACAGTGTGATGTCACCCCATGCTGCTGGTGTGCTCATCAACAGCAGA 240
 Db 408 GGTGAACCTTGGGAACAGTGTGATGTCACCCCATGCTGCTGGTGTGCTCATCAACAGCAGA 467
 Qy 241 TCCATGTCGCCCAACAGGACCATCCATGTCACCAAGTTCACCGACTGGAAGGCTACCTCATGA 300
 Db 468 TCCATGTCGCCCAACAGGACCATCCATGTCACCAAGTTCACCGACTGGAAGGCTACCTCATGA 527
 Qy 301 AACGGCTGTGGGCTCCAGACGCTTCCCGTGGATTTCCACATCAAGATGTTGGAGGCA 360
 Db 528 AACGGCTGTGGGCTCCAGACGCTTCCCGTGGATTTCCACATCAAGATGTTGGAGGCA 587
 Qy 361 TGAAGTACCCCTTTAGGACAGGCGATCGGCTGGAAGTGTGGAAGTCCCAAGTGTCTAC 420
 Db 588 TGAAGTACCCCTTTAGGACAGGCGATCGGCTGGAAGTGTGGAAGTCCCAAGTGTCTAC 647
 Qy 421 GCACTGCGATGCTGTGGTGGACACAGTAAATCGGGGCTCGCTTACCGCTCTCTACGAGG 480
 Db 648 GCACTGCGATGCTGTGGTGGACACAGTAAATCGGGGCTCGCTTACCGCTCTCTACGAGG 707
 Qy 481 ATGGTGACAGTACGACGACTTCTGGTGCCATGTGGAGCCCTTGATCCACCCAGTGG 540
 Db 708 ATGGTGACAGTACGACGACTTCTGGTGCCATGTGGAGCCCTTGATCCACCCAGTGG 767
 Qy 541 GTTGGTCAAGTGTGGGCCACCGGATCAAGATGTCAAGAGGCGAAGTGAATGAGGCC 600
 Db 768 GTTGGTCAAGTGTGGGCCACCGGATCAAGATGTCAAGAGGCGAAGTGAATGAGGCC 827
 Qy 601 ATCAACCCACCTTCCGGAGAGTCTACTGTGATGCGCTTCTTACCTCTTCAAGAGGTAC 660
 Db 828 ATCAACCCACCTTCCGGAGAGTCTACTGTGATGCGCTTCTTACCTCTTCAAGAGGTAC 887
 Qy 661 GAGCAGTCTACAGAGGCGGTTGGTTTGGAGAGGAGTGAAGTGGAGGCCATTGACC 720
 Db 888 GAGCAGTCTACAGAGGCGGTTGGTTTGGAGAGGAGTGAAGTGGAGGCCATTGACC 947
 Qy 721 CCCTGATCTGGGCAACATCTGCTGGCAACTGTCTGTAAGGTTCTCTGGATGATACC 780
 Db 948 CCCTGATCTGGGCAACATCTGCTGGCAACTGTCTGTAAGGTTCTCTGGATGATACC 1007
 Qy 781 TGATGATCTGTGGACGGGGGCCCTCCACAGATGGCTTGGACTGGTTCTGTCTACCATG 840
 Db 1008 TGATGATCTGTGGACGGGGGCCCTCCACAGATGGCTTGGACTGGTTCTGTCTACCATG 1067
 Qy 841 CCTTCTCCACGCCATCTTCCGGCCACCTTCTGTGTCAGAGAATGATGATGATGATCAC 900
 Db 1068 CCTTCTCCACGCCATCTTCCGGCCACCTTCTGTGTCAGAGAATGATGATGATGATCAC 1127
 Qy 901 CGCCAAAAGTTATGAGGCACACACTTCAACTGGGAGAACTACTTGGAGAGAGCAAGT 960
 Db 1128 CGCCAAAAGTTATGAGGCACACACTTCAACTGGGAGAACTACTTGGAGAGAGCAAGT 1187
 Qy 961 CGAAAAGCGCTCCATCGAGACTCTTTAATAGGATTTGCCCAACCATGCTTCAAGGTGG 1020
 Db 1188 CGAAAAGCGCTCCATCGAGACTCTTTAATAGGATTTGCCCAACCATGCTTCAAGGTGG 1247
 Qy 1021 GCATGAGCTGGAGCGGCTGATGATGAGAGCCCGGCTCATCTGTGTGGCAGGATGA 1080
 Db 1248 GCATGAGCTGGAGCGGCTGATGATGAGAGCCCGGCTCATCTGTGTGGCAGGATGA 1307
 Qy 1081 AACGAGTGTGATCGGCTCTCAGATCCACTTTGACCGGCTGGGACAGGATGACGCC 1140
 Db 1308 AACGAGTGTGATCGGCTCTCAGATCCACTTTGACCGGCTGGGACAGGATGACGCC 1367

Qy 1141 AGTGGTGGACTGCGAGTCCCCAGACATCTACCCCGTGGCTGGTGTGAGTCAACCGGCT 1200
 Db 1368 AGTGGTGGACTGCGAGTCCCCAGACATCTACCCCGTGGCTGGTGTGAGTCAACCGGCT 1427
 Qy 1201 ACCAGTCCAGGCTCTCTGTGGCGCAGAACCGGGCACAACCGCTGAAGGCGCAAGAGGCCA 1260
 Db 1428 ACCAGTCCAGGCTCTCTGTGGCGCAGAACCGGGCACAACCGCTGAAGGCGCAAGAGGCCA 1487
 Qy 1261 CAAGAAGAAAAGAAACAGTTTGGGAAGAAAAGAAAGAAATCCCGCCCACTAAGACGC 1320
 Db 1488 CAAGAAGAAAAGAAACAGTTTGGGAAGAAAAGAAAGAAATCCCGCCCACTAAGACGC 1547
 Qy 1321 GACCCCTCAGACAGGGGTCCAAGAACCCCTGTGGAGGACGACCCCTCAAGGTGCCAGGA 1380
 Db 1548 GACCCCTCAGACAGGGGTCCAAGAACCCCTGTGGAGGACGACCCCTCAAGGTGCCAGGA 1607
 Qy 1381 AGATCTCGTGGAGCCTGTTCTCTGGCGAGATCATTTGCTGTGGCGAGATCATTTGCTGTGGAGAGAGCATC 1440
 Db 1608 AGATCTCGTGGAGCCTGTTCTCTGGCGAGATCATTTGCTGTGGCGAGATCATTTGCTGTGGAGAGAGCATC 1667
 Qy 1441 TAGACGTGGCTCGCCCGACAAAGCTTCAAGTCCAGAGTGCCTGTCTCCGTCCAGAAACA 1500
 Db 1668 TAGACGTGGCTCGCCCGACAAAGCTTCAAGTCCAGAGTGCCTGTCTCCGTCCAGAAACA 1727
 Qy 1501 TCAAGCAGGAAAACAGACGACTGAGCCTTCTCTGCTCCAGCTGAGCTTCTAGCTGGAGCC 1560
 Db 1728 TCAAGCAGGAAAACAGACGACTGAGCCTTCTCTGCTCCAGCTGAGCTTCTAGCTGGAGCC 1787
 Qy 1561 AGCCACAGGCTTCTCTACCAACCAACAGCTGCTCCACCTGACTTTTGGCTTGGAGACTGA 1620
 Db 1788 AGCCACAGGCTTCTCTACCAACCAACAGCTGCTCCACCTGACTTTTGGCTTGGAGACTGA 1847
 Qy 1621 TCCTCTCTGTGTAATTTCTGCGCGTCTGTGAAGGCTGGAAGGCTGAGGACCTGCTGGG 1880
 Db 1848 TCCTCTCTGTGTAATTTCTGCGCGTCTGTGAAGGCTGGAAGGCTGAGGACCTGCTGGG 1907
 Qy 1681 GTCTCTGGGACCGCGCTGTTGCTGCTCCCTCCCTGTGAAAGGCTCTATATGACGGGC 1740
 Db 1908 GTCTCTGGGACCGCGCTGTTGCTGCTCCCTCCCTGTGAAAGGCTCTATATGACGGGC 1967
 Qy 1741 CGCTGAGGCCCCAAGAACTCGTGTGTAACCACTTTTCCAGCCAGAGTTTCCCAAAGCTG 1800
 Db 1968 CGCTGAGGCCCCAAGAACTCGTGTGTAACCACTTTTCCAGCCAGAGTTTCCCAAAGCTG 2027
 Qy 1801 GAACGCTAGCTGCTGCTCTTCTTAAGTGGCTCCCGCCGACCCCGCCAGCCCTCAG 1860
 Db 2028 GAACGCTAGCTGCTGCTCTTCTTAAAGTGGCTCCCGCCGACCCCGCCAGCCCTCAG 2087
 Qy 1861 TTGCCAGGGATGGGGCCCAACCACTGTCACTGTGGAATAACAGACAGTGAACCTGTCTG 1920
 Db 2088 TTGCCAGGGATGGGGCCCAACCACTGTCACTGTGGAATAACAGACAGTGAACCTGTCTG 2147
 Qy 1921 CCTGAACGAGTCATGTAAATTAAGTTCTAGACAGCTCTCTGAGCAGGATAAGGTCCCT 1980
 Db 2148 CCTGAACGAGTCATGTAAATTAAGTTCTAGACAGCTCTCTGAGCAGGATAAGGTCCCT 2207
 Qy 1981 GACAGTCAAGTTGTGTGGGGCAGCCTCTGCTCAAAAATTCACCAAGCAGAAATGCT 2040
 Db 2208 GACAGTCAAGTTGTGTGGGGCAGCCTCTGCTCAAAAATTCACCAAGCAGAAATGCT 2267
 Qy 2041 CTCAGCTCATGTGTTGGTCTCTGTCTCTCTAGCTCCCGAGGATGTTGGGGACCCAG 2100
 Db 2268 CTCAGCTCATGTGTTGGTCTCTGTCTCTCTAGCTCCCGAGGATGTTGGGGACCCAG 2327
 Qy 2101 CTTCGTCGCGCAGCTAAGAGCAGTGAACAGGATGTGGATTTTGGGACCTGTGTGTGG 2160
 Db 2328 CTTCGTCGCGCAGCTAAGAGCAGTGAACAGGATGTGGATTTTGGGACCTGTGTGTGG 2387
 Qy 2161 CTTTGAAGCTGCTTCTGTGTGTTGTGAGGACTGACTCCCAATTTCTTAAAGAAATGCCCC 2220
 Db 2388 CTTTGAAGCTGCTTCTGTGTGTTGTGAGGACTGACTCCCAATTTCTTAAAGAAATGCCCC 2447

QY 1141 AGTGGGTGAGCTCGAGTCCCAAGACATCTACCCCTCGGCTGGTGTGAGCTCACCGGCT 1200
Db 1368 AGTGGGTGAGCTCGAGTCCCAAGACATCTACCCCTCGGCTGGTGTGAGCTCACCGGCT 1427
QY 1201 ACCAGTCTCAGCTCTCTGTGGCGCAGAACCGGCACACCGCTGGAAGGCCAAGAGGCCA 1260
Db 1428 ACCAGTCTCAGCTCTCTGTGGCGCAGAACCGGCACACCGCTGGAAGGCCAAGAGGCCA 1487
QY 1261 CAAAGAAGAAAAAACAAGTTTGGAGAAAGAAAGAAATCCCGCCCACTAAGAGCC 1320
Db 1488 CAAAGAAGAAAAAACAAGTTTGGAGAAAGAAAGAAATCCCGCCCACTAAGAGCC 1547
QY 1321 GACCCCTCAGACAGGGGTCCAAAGAGCCCTCTGAGAGACAGCCCTCAGGCTGCCAGGA 1380
Db 1548 GACCCCTCAGACAGGGGTCCAAAGAGCCCTCTGAGAGACAGCCCTCAGGCTGCCAGGA 1607
QY 1381 AGATCTCGTGGAGCCCTGTTCTCTGGGAGATCATTTCTGTGGTGTGAAGGAAGAGCATC 1440
Db 1608 AGATCTCGTGGAGCCCTGTTCTCTGGGAGATCATTTCTGTGGTGTGAAGGAAGAGCATC 1667
QY 1441 TAGACGTGGCTCGCCGACAGGCTTCAAGTCAGAGTGCCTGTCTCCGTCGAGAAC 1500
Db 1668 TAGACGTGGCTCGCCGACAGGCTTCAAGTCAGAGTGCCTGTCTCCGTCGAGAAC 1727
QY 1501 TCAAGCAGGAAACAGAGCTGAGCTTCTGCTCCAGCCTGGCTTCTAGCTGGAAGCC 1560
Db 1728 TCAAGCAGGAAACAGAGCTGAGCTTCTGCTCCAGCCTGGCTTCTAGCTGGAAGCC 1787
QY 1561 AGCCAGCGTTTCTTACACACCAACCATGCTCCACCTGACTTTGGCTTGGAGACTGA 1620
Db 1788 AGCCAGCGTTTCTTACACACCAACCATGCTCCACCTGACTTTGGCTTGGAGACTGA 1847
QY 1621 TCCTCTCTGTAAATCTGCGCGTGTGTGAGGCTGAGCGTGGAGGACTGCTGGG 1680
Db 1848 TCCTCTCTGTAAATCTGCGCGTGTGTGAGGCTGAGCGTGGAGGACTGCTGGG 1907
QY 1681 GTCTCTGGAGCCGCTGTGCTTCTGCTCCCTGCTGGAAGGTCTATATGACGGGC 1740
Db 1908 GTCTCTGGAGCCGCTGTGCTTCTGCTCCCTGCTGGAAGGTCTATATGACGGGC 1967
QY 1741 CGCTGAGGCCCCAGAACTGCTGTGTGAACACCTTTTCAGCAGAGTTCCTCAAGCTG 1800
Db 1968 CGCTGAGGCCCCAGAACTGCTGTGTGAACACCTTTTCAGCAGAGTTCCTCAAGCTG 2027
QY 1801 GAAGCTAGTGTGCTGCTTCTTAAGATGGCTCCCGCCGACCGCCGACGCTCCAG 1860
Db 2028 GAAGCTAGTGTGCTGCTTCTTAAGATGGCTCCCGCCGACCGCCGACGCTCCAG 2087
QY 1861 TTGCCAGGATGGGGCCACCACTGTACACTGTGGAATACAAGACAGTGAATCTGTCTG 1920
Db 2088 TTGCCAGGATGGGGCCACCACTGTACACTGTGGAATACAAGACAGTGAATCTGTCTG 2147
QY 1921 CCTGACAGTCAATGTAATTAAGTTCTAGAGAGCTCTCTGAGCAGGATAGGTCCCT 1980
Db 2148 CCTGACAGTCAATGTAATTAAGTTCTAGAGAGCTCTCTGAGCAGGATAGGTCCCT 2207
QY 1981 GACAGTGAATGTGTGGTGGGGCAGCTCTGCTCAAAAATTCACCAAGCAGATGCT 2040
Db 2208 GACAGTGAATGTGTGGTGGGGCAGCTCTGCTCAAAAATTCACCAAGCAGATGCT 2267
QY 2041 CTAGCCCTCATGTGGTGGCTCTGCTCTCTCTAGCTCCCGAGGATGTGGGACCCAG 2100
Db 2268 CTAGCCCTCATGTGGTGGCTCTGCTCTCTCTAGCTCCCGAGGATGTGGGACCCAG 2327
QY 2101 CTGTCTCGCAGCTAAGAGCAGTACACAGGATGGATTTTGGCGACTGTGTGGTG 2160
Db 2328 CTGTCTCGCAGCTAAGAGCAGTACACAGGATGGATTTTGGCGACTGTGTGGTG 2387
QY 2161 CCTGAGCTCTTCTGTGTGTGTGAGGACTGACTCCCAATTCCTAAGGAAATGCCCC 2220
Db 2388 CCTGAGCTCTTCTGTGTGTGTGAGGACTGACTCCCAATTCCTAAGGAAATGCCCC 2447

QY 2221 GGGAGAGACATGGGAGGAAGATGGCTGAGTGTGACATTTTGGCTCTGTACCTGTCTCT 2280
Db 2448 GGGAGAGACATGGGAGGAAGATGGCTGAGTGTGACATTTTGGCTCTGTACCTGTCTCT 2507
QY 2281 GAAGCCCGCTAAAAATAATTCATCAAGATTCCTTTGTAGTTAAAGGGTCCAGTTCTGA 2340
Db 2508 GAAGCCCGCTAAAAATAATTCATCAAGATTCCTTTGTAGTTAAAGGGTCCAGTTCTGA 2567
QY 2341 CTGAGGCTCTAGAGAGCTGGCTTGTATCTTTTGGCTTTTGTCTTCTACCTAAATG 2400
Db 2568 TTGAGGCTCTAGAGAGCTGGCTTGTATCTTTTGGCTTTTGTCTTCTACCTAAATG 2627
QY 2401 AAGAAACATGCTGGAGGGCGGTGAACACAGAACCTCAAGACAGGATGACAGAGCT 2460
Db 2628 AAGAAACATGCTGGAGGGCGGTGAACACAGAACCTCAAGACAGGATGACAGAGCT 2687
QY 2461 GGAGGACACATCTAGCTGCCATTCGCAACCTCACTGGGCTCCCGAGACTCTGTGTGAGA 2520
Db 2688 GGAGGACACATCTAGCTGCCATTCGCAACCTCACTGGGCTCCCGAGACTCTGTGTGAGA 2747
QY 2521 AATTAAACCCCTGCTTGTCTTGAGAAAAA 2554
Db 2748 AATTAAACCCCTGCTTGTCTTGAGAAAAA 2781

RESULT 3

US-10-094-749-917
; Sequence 917, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 917
; LENGTH: 3188
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-917

Query Match 99.4%; Score 2538.8; DB 13; Length 3188;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2540; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AGCAGCTCCCACTCTATGACAGTGGGAGATGTGATGAAGGGATGAAGGTGGAGTGC 60
Db 647 AGCAGTCCCACTCTATGACAGTGGGAGATGTGATGAAGGGATGAAGGTGGAGTGC 706
QY 61 TCAACAGTGTGTGTGCTCTCCCGCGGGTGTACTGCGCTCTGTCTATCCAGACAG 120

Db 707 TCAACAGTGATCTGTGCTCCACAGCGGGTGACTGATCGCTCTGTGTCATCCAGACAG 766
Qy 121 CAGGGTATCGGGTCTGCTTCGGTATGAGGCTTTGAAATATGACCGCCATGACTTCT 180
Db 767 CAGGGTATCGGGTCTGCTTCGGTATGAGGCTTTGAAATATGACCGCCATGACTTCT 826
Qy 181 GGTGCAACCTGGGAACAGTGTGATCTCCACCCCATTTGGCTGGTGTGCCATCAACAGCAAGA 240
Db 827 GGTGCAACCTGGGAACAGTGTGATCTCCACCCCATTTGGCTGGTGTGCCATCAACAGCAAGA 886
Qy 241 TCCTAGTACCCCTTTAGGCAAGGATCCATGCCAAGTTTACCGACTGGAAGGGCTACCTCATGA 300
Db 887 TCCTAGTACCCCTTTAGGCAAGGATCCATGCCAAGTTTACCGACTGGAAGGGCTACCTCATGA 946
Qy 301 AACGECTGTGGGCTCCAGGACGCTTCCCGTGGATTTCCACATCAAGATGCTGGAGCA 360
Db 947 AACGECTGTGGGCTCCAGGACGCTTCCCGTGGATTTCCACATCAAGATGCTGGAGCA 1006
Qy 361 TGAAGTACCCCTTTAGGCAAGGATCCGCTGGAAGTGTGGACAAGTCCCAGGTGTAC 420
Db 1007 TGAAGTACCCCTTTAGGCAAGGATCCGCTGGAAGTGTGGACAAGTCCCAGGTGTAC 1066
Qy 421 GCATCTGCATGGCTGTGGTGGACACAGTATATCGGGGTGCGCTTACGGCTCCTCTACGAGG 480
Db 1067 GCATCTGCATGGCTGTGGTGGACACAGTATATCGGGGTGCGCTTACGGCTCCTCTACGAGG 1126
Qy 481 ATGGTGACAGTGCAGGACTTCTGTGTGGACATGTGGAGCGCCCTGATCCACCCAGTGG 540
Db 1127 ATGGTGACAGTGCAGGACTTCTGTGTGGACATGTGGAGCGCCCTGATCCACCCAGTGG 1186
Qy 541 GTTGGTCAAGCTGTGGGCGACGGATCAAGATGTGAGAGGGGATGAAGTGCATGGGCC 600
Db 1187 GTTGGTCAAGCTGTGGGCGACGGATCAAGATGTGAGAGGGGATGAAGTGCATGGGCC 1246
Qy 601 ATCACCCCACTTCCGGGAAGATCTACTGTGATCGCCGTTCCTTACCTTCAAGAGAGGTAC 660
Db 1247 ATCACCCCACTTCCGGGAAGATCTACTGTGATCGCCGTTCCTTACCTTCAAGAGAGGTAC 1306
Qy 661 GAGCAGTCTACAGAGGCGGTGTTTGAAGAGGGATGAAGTGGAGGCGCATTTGACC 720
Db 1307 GAGCAGTCTACAGAGGCGGTGTTTGAAGAGGGATGAAGTGGAGGCGCATTTGACC 1366
Qy 721 CCTGTAATCTGGCAACATCTCGGTGGCAACTCTGTGTAAGGTTCTCTGGATGATACC 1426
Db 1367 CCTGTAATCTGGCAACATCTCGGTGGCAACTCTGTGTAAGGTTCTCTGGATGATACC 1486
Qy 781 TGATGATCTGTGTGGACGGGGGCGCTCCACAGATGGCTTGGACTGGTTCTGCTACCATG 840
Db 1427 TGATGATCTGTGTGGACGGGGGCGCTCCACAGATGGCTTGGACTGGTTCTGCTACCATG 1486
Qy 841 CCTCTTCCACGCGCATCTTCCGGGCGACCTTCTGTGAGAGGAATGACATGAGCTCACAC 900
Db 1487 CCTCTTCCACGCGCATCTTCCGGGCGACCTTCTGTGAGAGGAATGACATGAGCTCACAC 1546
Qy 901 CGCAGAAAGGTTATGAGGACAGACTTTCATCTGGAGAACTACTTGGAGAGGACCAAGT 960
Db 1547 CGCAGAAAGGTTATGAGGACAGACTTTCATCTGGAGAACTACTTGGAGAGGACCAAGT 1606
Qy 961 CGAAAGCCGCTCATGAGACTTCTTAACTGATGATGCGGCAACCATGCTTCAAGGTGG 1020
Db 1607 CGAAAGCCGCTCATGAGACTTCTTAACTGATGATGCGGCAACCATGCTTCAAGGTGG 1666
Qy 1021 GCATGAGCTGGAGGCGGTGAGGACCTGATGGAGCCCGGCTCATCTGTGTGGCAGGTGA 1080
Db 1667 GCATGAGCTGGAGGCGGTGAGGACCTGATGGAGCCCGGCTCATCTGTGTGGCAGGTGA 1726
Qy 1081 AACGAGTGTGATCGGCTCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACGACC 1140
Db 1727 AACGAGTGTGATCGGCTCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACGACC 1786
Qy 1141 AGTGGGTGACCTGGAGTCCCAGACATCTACCCGCTCGGCTGTGTGAGTCTACCGGCT 1200

Db 1787 AGTGGGTGACCTGCGAGTCCCCAGACATCTACCCGCTCGGCTGTGTGAGCTACCGGCT 1846
Qy 1201 ACCAGCTCAGCCTCTCTGTGCGCGAGAACCGGCCACACCCGCTGAAGGCCAAAAGAGGCCA 1260
Db 1847 ACCAGCTCAGCCTCTCTGTGCGCGAGAACCGGCCACACCCGCTGAAGGTCAAAGAGGCCA 1906
Qy 1261 CAAAGCAAGAAAAGAAACAGTTTGGGAGAAAAGAAAAGAAATCCCGCCCACTAAGAGCC 1320
Db 1907 CAAAGCAAGAAAAGAAACAGTTTGGGAGAAAAGAAAAGAAATCCCGCCCACTAAGAGCC 1966
Qy 1321 GACCCCTCAGACAGGGGTCCAAAGAGCCCTGTCTGGAGAGCAGCCCTCAGGGTGCACAGA 1380
Db 1967 GACCCCTCAGACAGGGGTCCAAAGAGCCCTGTCTGGAGAGCAGCCCTCAGGGTGCACAGA 2026
Qy 1381 AGATCTCTCGAGGAGCTTCTCTGCGAGATCATTTGCTGTGCTGTGAAGAAAGAGCATC 1440
Db 2027 AGATCTCTCGAGGAGCTTCTCTGCGAGATCATTTGCTGTGCTGTGAAGAAAGAGCATC 2086
Qy 1441 TAGAGCTGGCCTCGCCGACAAAGCTTCAAGTCCAGAGCTGCCCTGTCTCCGTCAGAAACA 1500
Db 2087 TAGAGCTGGCCTCGCCGACAAAGCTTCAAGTCCAGAGCTGCCCTGTCTCCGTCAGAAACA 2146
Qy 1501 TCAAGCAGAAAACAGACGACTGAGCCTTCTGCTCCAGCCTGGCTTCTAGCTGGAAGCC 1560
Db 2147 TCAAGCAGAAAACAGACGACTGAGCCTTCTGCTCCAGCCTGGCTTCTAGCTGGAAGCC 2206
Qy 1561 AGCCAGAGTTTCTTACCAACACCATGCTCCACCTGACTTTGGCTTGGAGACTGA 1620
Db 2207 AGCCAGAGTTTCTTACCAACACCATGCTCCACCTGACTTTGGCTTGGAGACTGA 2266
Qy 1621 TCCTCTCTGTGTAATTTCTGCCGCTGTGTGAAGGCTGACGCTGAGGACCTGTCTGGG 1680
Db 2267 TCCTCTCTGTGTAATTTCTGCCGCTGTGTGAAGGCTGACGCTGAGGACCTGTCTGGG 2326
Qy 1681 GTCTCTGGGACCGCCTGTGTCTGCGCTGCCCTGTGCGGAAAGGCTTATATGACGGGC 1740
Db 2327 GTCTCTGGGACCGCCTGTGTCTGCGCTGCCCTGTGCGGAAAGGCTTATATGACGGGC 2386
Qy 1741 CGCTGAGGGCCCCAGAACTCTGTGTAACACACTTTTCCAGCCAGAGTTTCCCAAGCTG 1800
Db 2387 CGCTGAGGGCCCCAGAACTCTGTGTAACACACTTTTCCAGCCAGAGTTTCCCAAGCTG 2446
Qy 1801 GAACGCTAGCTGCTGTCTTCTTAAAGTGGCTCCCCCGACCCGACCGCCCTCAG 1860
Db 2447 GAACGCTAGCTGCTGTCTTCTTAAAGTGGCTCCCCCGACCCGACCGCCCTCAG 2506
Qy 1861 TTGCCAGGGATGGGGCCACCACTGTCACTGTGGATACAAGACAGTGAACCTGTCTG 1920
Db 2507 TTGCCAGGGATGGGGCCACCACTGTCACTGTGGATACAAGACAGTGAACCTGTCTG 2566
Qy 1921 CCTGAACGAGTCTATGTAATTTAAAGTTCTAGACAGCTCTCTGAGCAGGATAAGGTCCT 1980
Db 2567 CCTGAACGAGTCTATGTAATTTAAAGTTCTAGACAGCTCTCTGAGCAGGATAAGGTCCT 2626
Qy 1981 GACAGTGAAGTGTGGTGGGGCAGCCTTCTGCTCAAAAATTCACCAAGCAGAAATGCCT 2040
Db 2627 GACAGTGAAGTGTGGTGGGGCAGCCTTCTGCTCAAAAATTCACCAAGCAGAAATGCCT 2686
Qy 2041 CTGAGCTCATGTGTGCTCTCTGCTCTCTAGCTCCCCAGGATGTGGGGACCCAG 2100
Db 2687 CTGAGCTCATGTGTGCTCTCTGCTCTCTAGCTCCCCAGGATGTGGGGACCCAG 2746
Qy 2101 CTTGTCTCGGACCTAAGAAAGCAGTGAACAGGATGGATTTTGGCGACCTGTGTGGTG 2160
Db 2747 CTTGTCTCGGACCTAAGAAAGCAGTGAACAGGATGGATTTTGGCGACCTGTGTGGTG 2806
Qy 2161 CTTGAGCTGCTTCTGTGTGTGTGAGGACTGACTCCCAATTTCCCTAAAGAAATGCCCC 2220
Db 2807 CTTGAGCTGCTTCTGTGTGTGTGAGGACTGACTCCCAATTTCCCTAAAGAAATGCCCC 2866
Qy 2221 GGGAGGACATTTGGGAGGAGATGGCCTGAGTGTGCACTTTGGCTCTGCTACCTGCTCT 2280
Db 2867 GGGAGGACATTTGGGAGGAGATGGCCTGAGTGTGCACTTTGGCTCTGCTACCTGCTCT 2926

QY 2281 GAAGCCCGCTAAATAATATCATCCAGATTCTTTGTAGTTAAAGGGTCCAGTTCTGA 2340
 DB 2927 GAAGCCCGCTAAATAATATCATCCAGATTCTTTGTAGTTAAAGGGTCCAGTTCTGA 2986
 QY 2341 CTGAGAGCTCTAGAGAGCTGGCTGTATGTTCTTTTGGGCTTTTGTCTTACCTAAATG 2400
 DB 2987 TTGAGAGCTCTAGAGAGCTGGCTGTATGTTCTTTTGGGCTTTTGTCTTACCTAAATG 3046
 QY 2401 AAGAAACCATGCTCGAGGGGCGGTGAACACAGAACCCCTCAAGACAAGGATGACAGCT 2460
 DB 3047 AAGAAACCATGCTCGAGGGGCGGTGAACACAGAACCCCTCAAGACAAGGATGACAGCT 3106
 QY 2461 GGAGGACATCTAGCTGACCTGCAACCTCACTGGGCTCCCGAGACTCTGTGTGAGA 2520
 DB 3107 GGAGGACATCTAGCTGACCTGCAACCTCACTGGGCTCCCGAGACTCTGTGTGAGA 2520
 QY 2521 AATTAAACCCCTGCTTGTGTTG 2542
 DB 3167 AATTAAACCCCTGCTTGTGTTG 3188

RESULT 4

US-10-264-237-419
 ; Sequence 419, Application US/10264237
 ; Publication No. US2004009491A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Birse et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PA131P1
 ; CURRENT APPLICATION NUMBER: US/10/264,237
 ; PRIOR FILING DATE: 2002-10-04
 ; PRIOR APPLICATION NUMBER: PCT/US01/16450
 ; PRIOR FILING DATE: 2001-05-18
 ; PRIOR APPLICATION NUMBER: US 60/205,515
 ; PRIOR FILING DATE: 2000-05-19
 ; NUMBER OF SEQ ID NOS: 2876
 ; SOFTWARE: PatentIn Ver. 3.1
 ; SEQ ID NO 419
 ; LENGTH: 2663
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (660)..(660)
 ; OTHER INFORMATION: n equals a.t.g, or c
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (775)..(775)
 ; OTHER INFORMATION: n equals a.t.g, or c
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1248)..(1248)
 ; OTHER INFORMATION: n equals a.t.g, or c
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1824)..(1824)
 ; OTHER INFORMATION: n equals a.t.g, or c
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (2642)..(2642)
 ; OTHER INFORMATION: n equals a.t.g, or c
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (2648)..(2648)
 ; OTHER INFORMATION: n equals a.t.g, or c
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (2662)..(2662)
 ; OTHER INFORMATION: n equals a.t.g, or c
 ; US-10-264-237-419

Query Match

98.4%; Score 2514.4; DB 12; Length 2663;

Best Local Similarity 99.7%; Pred. No. 0;
 Matches 2537; Conservative 1; Mismatches 5; Indels 2; Gaps 2;
 QY 1 AGCAGTCCCACTCTATGACAGTGGGAGGATGTGATGAAGGATGAAGTGGAGTGC 60
 DB 72 AGCAGTCCCACTCTATGACAGTGGGAGGATGTGATGAAGGATGAAGTGGAGTGC 131
 QY 61 TCAACAGTATGTGTCTCTCCAGCCGGGTGTACTGGATCGCTCTGTATCCAGACAG 120
 DB 132 TCAACAGTATGTGTCTCTCCAGCCGGGTGTACTGGATCGCTCTGTATCCAGACAG 191
 QY 121 CAGGATATCGGCTGCTCTCGGTATGAAGCTTTGAAATGAGCCAGCCATGACTTCT 180
 DB 192 CAGGATATCGGCTGCTCTCGGTATGAAGCTTTGAAATGAGCCAGCCATGACTTCT 251
 QY 181 GGTGCAACTGGGAAACAGTGTGATGTCACCCCAATGGCTGGTGTGCCATCAACAGCAAGA 240
 DB 252 GGTGCAACTGGGAAACAGTGTGATGTCACCCCAATGGCTGGTGTGCCATCAACAGCAAGA 311
 QY 241 TCCTAGTGGCCCAAGGACCATCCATGCGCAAGTTACCGACTGGAGGCTACCTATGA 300
 DB 312 TCCTAGTGGCCCAAGGACCATCCATGCGCAAGTTACCGACTGGAGGCTACCTATGA 371
 QY 301 AACGGCTGGTGGCTCCAGGACGCTTCCCGTGGATTTCCACATCAAGATGGTGGAGACCA 360
 DB 372 AACGGCTGGTGGCTCCAGGACGCTTCCCGTGGATTTCCACATCAAGATGGTGGAGACCA 431
 QY 361 TGAAGTACCCCTTTAGCAGGAGGATGCGGCTGGAAGTGGTGGACAACTCCAGGTGTAC 420
 DB 432 TGAAGTACCCCTTTAGCAGGAGGATGCGGCTGGAAGTGGTGGACAACTCCAGGTGTAC 491
 QY 421 GCATCTCGATGGCTGTGGTGGACACAGTAAATCGGGGTGCGCTTACGGCTCTCTTACGAGG 480
 DB 492 GCATCTCGATGGCTGTGGTGGACACAGTAAATCGGGGTGCGCTTACGGCTCTCTTACGAGG 551
 QY 481 ATGTGACAGTACGACGACTTCTGTGTGTCACATGTGGAGCCCGCTGTATCCACCCAGTGG 540
 DB 552 ATGTGACAGTACGACGACTTCTGTGTGTCACATGTGGAGCCCGCTGTATCCACCCAGTGG 611
 QY 541 GTTGGTCAAGACGCTGGGCGACGGATCAAGATGTGTCAGAGGCGAAGTGCATGGCCCC 600
 DB 612 GTTGGTCAAGACGCTGGGCGACGGATCAAGATGTGTCAGAGGCGAAGTGCATGGCCCC 671
 QY 601 ATCAACCCCACTTCCGGAAGATCTACTGTGATGCCGTTTCTTACCTCTTCAAGAAGGTAC 660
 DB 672 ATCAACCCCACTTCCGGAAGATCTACTGTGATGCCGTTTCTTACCTCTTCAAGAAGGTAC 731
 QY 661 GAGCAGTCTACACAGAGGCGGTTGGTTGAGGAAGGATGAAGCTGGAGGCCATGACCC 720
 DB 732 GAGCAGTCTACACAGAGGCGGTTGGTTGAGGAAGGATGAAGCTGGAGGCCATGACCC 791
 QY 721 CCCTGAATCTGGGCAACATCTGGTGGCAACTGTCTGTAAGGTTCTCTGGATGGATACC 780
 DB 792 CCCTGAATCTGGGCAACATCTGGTGGCAACTGTCTGTAAGGTTCTCTGGATGGATACC 851
 QY 781 TGATGATCTGTGTGGACGGGGGCGCTTCCACAGATGGCTTGGATCTGGTCTGTACATG 840
 DB 852 TGATGATCTGTGTGGACGGGGGCGCTTCCACAGATGGCTTGGATCTGGTCTGTACATG 911
 QY 841 CCTCTTCCACAGCCTTCTTCCCGGCCACCTTCTGTGTCAGAGGATGACATTTGAGCTCACAC 900
 DB 912 CCTCTTCCACAGCCTTCTTCCCGGCCACCTTCTGTGTCAGAGGATGACATTTGAGCTCACAC 971
 QY 901 CGCCAAAAGGTTATGAGGACAGACTTTCACTGGGAGAACTACTTTGGAGAGACCAAGT 960
 DB 972 CGCCAAAAGGTTATGAGGACAGACTTTCACTGGGAGAACTACTTTGGAGAGACCAAGT 1031
 QY 961 CGAAAGCCGCTCCATCGAGACTCTTTAAAGTGGTGGGCGGCGGCTTCAAGGTG 1020
 DB 1032 CGAAAGCCG--TCCATCGAGACTCTTTAAAGTGGTGGGCGGCGGCTTCAAGGTG 1090
 QY 1021 GCATGAAGCTGGAGGCGCTGGACCTGATGGAGCCCGGCTCATCTGTGTGGCCACCGTGA 1080

Db 1091 GCATGAAGCTGGAGGCGGTGACCTGATGAGCCCCGGGCTCATCTGTGTGGCCACGGTGA 1150
QY 1081 AACGAGTGTGATCGGCTCCCTCAGCATCCATTTGACGGCTGGACACGAGTACGACC 1140
Db 1151 AACGAGTGTGATCGGCTCCCTCAGCATCCATTTGACGGCTGGACACGAGTACGACC 1210
QY 1141 AGTGGTGAAGTCCGAGTCCCGAGCATCTACCCGGT - CGGCTGGTGTGAGCTACCCGGC 1199
Db 1211 AGTGGTGAAGTCCGAGTCCCGAGCATCTACCCGGTTCGCTGTGAGCTACCCGGC 1270
QY 1200 TACCAGCTCCAGCTCTGTGGCCGAGAACCGGCCACACCCGCTGAAGGCCMAAGAGGCC 1259
Db 1271 TACCAGCTCCAGCTCTGTGGCCGAGAACCGGCCACACCCGCTGAAGGCCMAAGAGGCC 1330
QY 1260 ACAAGAAGAAAAGAAACAGTTTGGGAAGAAAAGAAAGAAATCCCGCCCACTAAGAGC 1319
Db 1331 ACAAGAAGAAAAGAAACAGTTTGGGAAGAAAAGAAAGAAATCCCGCCCACTAAGAGC 1390
QY 1320 CGACCCCTCAGACAGGGGTCCAAAGAGCCCTGTGGAGGACGACCCCTCAGGGTCCAGG 1379
Db 1391 CGACCCCTCAGACAGGGGTCCAAAGAGCCCTGTGGAGGACGACCCCTCAGGGTCCAGG 1450
QY 1380 AAGATCTCTCGGAGCGCTCTCTCGCGAGATCATGTGTCGCTGTGAGAGAGAGCAT 1439
Db 1451 AAGATCTCTCGGAGCGCTCTCTCGCGAGATCATGTGTCGCTGTGAGAGAGAGCAT 1510
QY 1440 CTAGAGCTGGCCCTCGCCCGACAGGCTTCAAGTCCAGAGCTGCCTGTCTCCGTCGAGAAC 1499
Db 1511 CTAGAGCTGGCCCTCGCCCGACAGGCTTCAAGTCCAGAGCTGCCTGTCTCCGTCGAGAAC 1570
QY 1500 ATCAAGCAGAAACAGACAGCTGAGGCTTCGCTCCAGCTCGAGCTGCTTCTAGCTGGAAGC 1559
Db 1571 ATCAAGCAGAAACAGACAGCTGAGGCTTCGCTCCAGCTCGAGCTGCTTCTAGCTGGAAGC 1630
QY 1560 CAGCCAGAGGTTTCTTACACACACACACATCCCTCCAGCTGACCTTGGCTTGGAGACTG 1619
Db 1631 CAGCCAGAGGTTTCTTACACACACACACATCCCTCCAGCTGACCTTGGCTTGGAGACTG 1690
QY 1620 ATCTCTCTGTGTAAATCTGCCCGGTGTGTGAAGGCTGGAAGCTGGAGGACCTGTCTG 1679
Db 1691 ATCTCTCTGTGTAAATCTGCCCGGTGTGTGAAGGCTGGAAGCTGGAGGACCTGTCTG 1750
QY 1680 GGTCTCTGGGACCGCTGTGTCTTCTGCTCCCTCCCTGTGGAAGGCTTATATGACGGG 1739
Db 1751 GGTCTCTGGGACCGCTGTGTCTTCTGCTCCCTCCCTGTGGAAGGCTTATATGACGGG 1810
QY 1740 CGGCTTGAGGCCCCAGAACTCTGCTGTGAACCACTTTTCCAGCCAGAGTTCCCAAGCT 1799
Db 1811 CGGCTTGAGGCCCCAGAACTCTGCTGTGAACCACTTTTCCAGCCAGAGTTCCCAAGCT 1870
QY 1800 GGAACGCTAGTCTGCTCTTCTTAAGATGGCTTCCCGCCGACCCCGCCACGGCCCTCA 1859
Db 1871 GGAACGCTAGTCTGCTCTTCTTAAGATGGCTTCCCGCCGACCCCGCCACGGCCCTCA 1930
QY 1860 GTTGCCAGGATGGGGCCACCACTGTCACTGTGGATACAGACAGTGAACCTGTCT 1919
Db 1931 GTTGCCAGGATGGGGCCACCACTGTCACTGTGGATACAGACAGTGAACCTGTCT 1990
QY 1920 GCTGAAACGAGTCATGTAATTAAGTTCTAGACAGCTCTCTGAGCAGATAAGTCTCCC 1979
Db 1991 GCTGAAACGAGTCATGTAATTAAGTTCTAGACAGCTCTCTGAGCAGATAAGTCTCCC 2050
QY 1980 TGACAGTGAATGTGTGTGGGGGAGCGCTCTGCTTCAAAATTCACCAAGAGATGCC 2039
Db 2051 TGACAGTGAATGTGTGTGGGGGAGCGCTCTGCTTCAAAATTCACCAAGAGATGCC 2110
QY 2040 TCTCAGCCTCATGTGTGTGTCTCTCTCTCTAGCTCCCGAGGATGTGGGACCCA 2099
Db 2111 TCTCAGCCTCATGTGTGTGTCTCTCTCTCTAGCTCCCGAGGATGTGGGACCCA 2170
QY 2100 GCTTGTCTGGCAGCTAAGAACAGTACAGGATGTGGATTTTGGGACCTGTGTGTG 2159
Db 2171 GCTTGTCTGGCAGCTAAGAACAGTACAGGATGTGGATTTTGGGACCTGTGTGTG 2230

QY 2160 GCCTTGAGCTGCTTTCTGTGTGTGTGTGAGGACTGACTCCCATTTCTTAAAGAAATGCCCC 2219
Db 2231 GCCTTGAGCTGCTTTCTGTGTGTGTGTGAGGACTGACTCCCATTTCTTAAAGAAATGCCCC 2290
QY 2220 CGGGGAGACATTTGGGAGGAAGATGGCTCAGTGTGCACCTTTGGCTCTGTACTGTCTCC 2279
Db 2291 CGGGGAGACATTTGGGAGGAAGATGGCTCAGTGTGCACCTTTGGCTCTGTACTGTCTCC 2350
QY 2280 TGAAGCCCGCTTAAATAATTTCAATCAAGATTCCTTTGTAGTTAAAGGGTCCAGTCTG 2339
Db 2351 TGAAGCCCGCTTAAATAATTTCAATCAAGATTCCTTTGTAGTTAAAGGGTCCAGTCTG 2410
QY 2340 ACTGAGCCCTTAGAGAGCTGGCTGTATGTCTTTTGGCTTTTCTTACCTAAAT 2399
Db 2411 ACTGAGCCCTTAGAGAGCTGGCTGTATGTCTTTTGGCTTTTCTTACCTAAAT 2470
QY 2400 GAAGAAACCATGTCTGGAGGGCGGTGAACACAGAACCTTCAAGACAAGGATCAGAGC 2459
Db 2471 GAAGAAACCATGTCTGGAGGGCGGTGAACACAGAACCTTCAAGACAAGGATCAGAGC 2530
QY 2460 TGGAGGACACATCTAGCTGCATTGCAACTCACTGGCTCCCGAGACTCTGTGTGTGAG 2519
Db 2531 TGGAGGACACATCTAGCTGCATTGCAACTCACTGGCTCCCGAGACTCTGTGTGTGAG 2590
QY 2520 AAATTTAAACCCCTGTCTTGTCTGAG 2544
Db 2591 AAATTTAAACCCCTGTCTTGTCTGAG 2615

RESULT 5

US-10-108-260A-1658
; Sequence 1658, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cdna
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1658
; LENGTH: 2432
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1658

Query Match 48.1%; Score 1227.8; DB 12; Length 2432;

Best Local Similarity 99.4%; Pred. No. 0;

Matches 1232; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 AGCACGTCCCACTCTATGACCACTGGAGGATGTGATGAAGGATGAAGGTGGAGTGC 60
Db 719 AGCACGTCCCACTCTATGACCACTGGAGGATGTGATGAAGGATGAAGGTGGAGTGC 778
QY 61 TCAACAGTGAATGTGTGCTCCCGAGCCGGTGTACTGTGATCGCTCTGTCTATCAGACAG 120
Db 779 TCAACAGTGAATGTGTGCTCCCGAGCCGGTGTACTGTGATCGCTCTGTCTATCAGACAG 838
QY 121 CAGGGTATCGGTGCTGCTTCGGTATGAAGGCTTTGAAATGACGCCAGCCATGCTTCT 180
Db 839 CAGGGTATCGGTGCTGCTTCGGTATGAAGGCTTTGAAATGACGCCAGCCATGCTTCT 898
QY 181 GGTGCAACCTTGGGAAACAGTGGATGTCCACCCCATTTGGCTGGTGTGTCATCAACAGCAAG 240
Db 899 GGTGCAACCTTGGGAAACAGTGGATGTCCACCCCATTTGGCTGGTGTGTCATCAACAGCAAG 958
QY 241 TCTAGTGTCCCGACCGGACCATTCATGCGCAAGTTCACCGACTGGAAGGCTTACCTCATGA 300
Db 959 TCTAGTGTCCCGACCGGACCATTCATGCGCAAGTTCACCGACTGGAAGGCTTACCTCATGA 1018
QY 301 AACGGCTGGTGGGCTCCAGGACGCTTCCCGTGGATTTTCCACATCAAGATGGTGGAGAGCA 360


```

; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70683
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-70683

```

```

Query Match      20.0%; Score 512.2; DB 13; Length 513;
Best Local Similarity 99.6%; Pred. No. 1e-147;
Matches 511; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1439 TCTAGACGTGGCTCGCCGCGACAGAGCTTCAAGTCCAGAGCTGCTGCTCCGTCGAGAA 1498
Db 1 TCTAGACGTGGCTCGCCGCGACAGAGCTTCAAGTCCAGAGCTGCTGCTCCGTCGAGAA 60

QY 1499 CATCAAGCAGAGAAACAGACGACTGAGCTTCTGCTCCAGCTGCTTCAAGTGGAG 1558
Db 61 CATCAAGCAGAGAAACAGACGACTGAGCTTCTGCTCCAGCTGCTTCAAGTGGAG 120

QY 1559 CCAGCCAGAGGTTTCTTCTACCAACACCATGCTCCAGCTGCTTGGCTTGGAGACT 1618
Db 121 CCAGCCAGAGGTTTCTTCTACCAACACCATGCTCCAGCTGCTTGGCTTGGAGACT 180

QY 1619 GATCCTCTCTGTAAATTTCTGCTCCGCTGCTGTAAGGCTGAGAGCTGAGAGCTGCTG 1678
Db 181 GATCCTCTCTGTAAATTTCTGCTCCGCTGCTGTAAGGCTGAGAGCTGAGAGCTGCTG 240

QY 1679 GGGTCTCTCTGGACCCGCTGTTGCTTCTGCTCCGCTGCTGTAAGGCTGAGAGCTGAGAG 1738
Db 241 GGGTCTCTCTGGACCCGCTGTTGCTTCTGCTCCGCTGCTGTAAGGCTGAGAGCTGAGAG 300

QY 1739 GCGGCTGAGGCGCCAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCCAAAGC 1798
Db 301 GCGGCTGAGGCGCCAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCCAAAGC 360

QY 1799 TGAAGCTAGTGGCTGCTGCTTCTTAAAGTGGCTTCCGCTGCTGTAAGGCTGAGAGCTGAGAG 1858
Db 361 TGAAGCTAGTGGCTGCTGCTTCTTAAAGTGGCTTCCGCTGCTGTAAGGCTGAGAGCTGAGAG 420

QY 1859 AGTTGCCAGGATGGGCGCCACCACTGCTCACTGTGGAATCAAGACAGTGAATCTGTC 1918
Db 421 AGTTGCCAGGATGGGCGCCACCACTGCTCACTGTGGAATCAAGACAGTGAATCTGTC 480

QY 1919 TGCCTGAACGAGTCAATGTAATTAAGTTCTAGA 1951
Db 481 TGCCTGAACGAGTCAATGTAATTAAGTTCTAGA 513

```

```

RESULT 8
US-10-027-632-70682
; Sequence 70682, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome

```

```

; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70682
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-70682

```

```

Query Match      20.0%; Score 512.2; DB 14; Length 513;
Best Local Similarity 99.6%; Pred. No. 1e-147;
Matches 511; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1439 TCTAGACGTGGCTCGCCGCGACAGAGCTTCAAGTCCAGAGCTGCTGCTCCGTCGAGAA 1498
Db 1 TCTAGACGTGGCTCGCCGCGACAGAGCTTCAAGTCCAGAGCTGCTGCTCCGTCGAGAA 60

QY 1499 CATCAAGCAGAGAAACAGACGACTGAGCTTCTGCTCCAGCTGCTTCAAGTGGAG 1558
Db 61 CATCAAGCAGAGAAACAGACGACTGAGCTTCTGCTCCAGCTGCTTCAAGTGGAG 120

QY 1559 CCAGCCAGAGGTTTCTTCTACCAACACCATGCTCCAGCTGCTTGGCTTGGAGACT 1618
Db 121 CCAGCCAGAGGTTTCTTCTACCAACACCATGCTCCAGCTGCTTGGCTTGGAGACT 180

QY 1619 GATCCTCTCTGTAAATTTCTGCTCCGCTGCTGTAAGGCTGAGAGCTGAGAGCTGCTG 1678
Db 181 GATCCTCTCTGTAAATTTCTGCTCCGCTGCTGTAAGGCTGAGAGCTGAGAGCTGCTG 240

QY 1679 GGGTCTCTCTGGACCCGCTGTTGCTTCTGCTCCGCTGCTGTAAGGCTGAGAGCTGAGAG 1738
Db 241 GGGTCTCTCTGGACCCGCTGTTGCTTCTGCTCCGCTGCTGTAAGGCTGAGAGCTGAGAG 300

QY 1739 GCGGCTGAGGCGCCAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCCAAAGC 1798
Db 301 GCGGCTGAGGCGCCAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCCAAAGC 360

QY 1799 TGAAGCTAGTGGCTGCTGCTTCTTAAAGTGGCTTCCGCTGCTGTAAGGCTGAGAGCTGAGAG 1858
Db 361 TGAAGCTAGTGGCTGCTGCTTCTTAAAGTGGCTTCCGCTGCTGTAAGGCTGAGAGCTGAGAG 420

QY 1859 AGTTGCCAGGATGGGCGCCACCACTGCTCACTGTGGAATCAAGACAGTGAATCTGTC 1918
Db 421 AGTTGCCAGGATGGGCGCCACCACTGCTCACTGTGGAATCAAGACAGTGAATCTGTC 480

QY 1919 TGCCTGAACGAGTCAATGTAATTAAGTTCTAGA 1951
Db 481 TGCCTGAACGAGTCAATGTAATTAAGTTCTAGA 513

```

```

RESULT 9
US-10-027-632-70683
; Sequence 70683, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome

```

FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-08-09
PRIOR APPLICATION NUMBER: US 60/146,002
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 70683
LENGTH: 513
TYPE: DNA
ORGANISM: Human
US-10-027-632-70683

Query Match 20.0%; Score 512.2; DB 14; Length 513;
Best Local Similarity 99.6%; Pred. No. 1e-147; 0; Indels 0; Gaps 0;
Matches 511; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1439 TCTAGACGTGGCTCCGCCGCAAGGCTTCAAGTCAGAGCTGCCTGTCTCCGTCGAGAA 1498
Db 1 TCTAGACGTGGCTCCGCCGCAAGGCTTCAAGTCAGAGCTGCCTGTCTCCGTCGAGAA 60
QY 1499 CATCAAGCAGGAAACAGACGCTGAGCTTCTGCTCCAGCTGGCTTCTAGCTGGAG 1558
Db 61 CATCAAGCAGGAAACAGACGCTGAGCTTCTGCTCCAGCTGGCTTCTAGCTGGAG 120
QY 1559 CCAGCCGAGCTTCTCTACCAACACCACTGCTCCACCTGACTTTGGCTTGGAGACT 1618
Db 121 CCAGCCGAGCTTCTCTACCAACACCACTGCTCCACCTGACTTTGGCTTGGAGACT 180
QY 1619 GATCCTCTGTGTAAATTCGCGCGTCTGTGAAGCTGACGGTGGAGACCTGTCTG 1678
Db 181 GATCCTCTGTGTAAATTCGCGCGTCTGTGAAGCTGACGGTGGAGACCTGTCTG 240
QY 1679 GGGTCTCTGGACCCGCTGTGTCTGCTCCCTCCCTGCGAAAGTCTATATGACGG 1738
Db 241 GGGTCTCTGGACCCGCTGTGTCTGCTCCCTCCCTGCGAAAGTCTATATGACGG 300
QY 1739 GCGGCTGAGGCCCCAGAACTGCTGTGAACCACTTTTCCAGCCAGAGTTCCCAAGC 1798
Db 301 GCGGCTGAGGCCCCAGAACTGCTGTGAACCACTTTTCCAGCCAGAGTTCCCAAGC 360
QY 1799 TGAACGCTAGCTGCTGTCTTCTTAAGTGGCTCCCGCCGACCCGCGCCGCTC 1858
Db 361 TGAACGCTAGCTGCTGTCTTCTTAAGTGGCTCCCGCCGACCCGCGCCGCTC 420
QY 1859 AGTTGCCAGGATGGGCCCCACCACTGTACACTGTGGAATACAGACAGTCAACTGTCTG 1918
Db 421 AGTTGCCAGGATGGGCCCCACCACTGTACACTGTGGAATACAGACAGTCAACTGTCTG 480
QY 1919 TGCCTGAACGATCATGTAAATTAAGTTCTAGA 1951
Db 481 TGCCTGAACGATCATGTAAATTAAGTTCTAGA 513

RESULT 10
US-10-027-632-293974
Sequence 293974, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 293974
LENGTH: 472
TYPE: DNA
ORGANISM: Human
US-10-027-632-293974

Query Match 18.4%; Score 471.2; DB 13; Length 472;
Best Local Similarity 99.6%; Pred. No. 4.9e-135;
Matches 470; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1480 TGCTGTCTCCGTCGAGAACATCAAGCAGGAAACAGACGCTGAGCTTCTCCCTCCAG 1539
Db 1 TGCTGTCTCCGTCGAGAACATCAAGCAGGAAACAGACGCTGAGCTTCTCCCTCCAG 60
QY 1540 CTTGGCTTCTAGCTGGAAGCCAGCCAGCGTTTCTTACCACCAACCACTGCTCCACC 1599
Db 61 CTTGGCTTCTAGCTGGAAGCCAGCCAGCGTTTCTTACCACCAACCACTGCTCCACC 120
QY 1600 TGACTTTGGCTTGGAGACTGATCTCTGTGTAAATTCGCGCGTGTGTGAAGCTG 1659
Db 121 TGACTTTGGCTTGGAGACTGATCTCTGTGTAAATTCGCGCGTGTGTGAAGCTG 180
QY 1660 GACGCTGGAGGACCTCTCTGGGCTCTCTGGGACCCGCTGTGTCTGCTCCCTCTG 1719
Db 181 GACGCTGGAGGACCTCTCTGGGCTCTCTGGGACCCGCTGTGTCTGCTCCCTCTG 240
QY 1720 GGAAGGCTCTATATGACGGGCGCTGTGAGGCGCCAGAACTGCTGTGTGAACCACTTTTC 1779
Db 241 GGAAGGCTCTATATGACGGGCGCTGTGAGGCGCCAGAACTGCTGTGTGAACCACTTTTC 300
QY 1780 CAGCCAGAGTTCCCAAGCTGGAACGCTAGCTGCTTCTTCCCTTAAGATGGCTCCCC 1839
Db 301 CAGCCAGAGTTCCCAAGCTGGAACGCTAGCTGCTTCTTCCCTTAAGATGGCTCCCC 360
QY 1840 CCGACCCGCGACGGCCCTCAGTTGCGAGGATGGGCGCCACCACTGTACACTGTGAATA 1899
Db 361 CCGACCCGCGACGGCCCTCAGTTGCGAGGATGGGCGCCACCACTGTACACTGTGAATA 420
QY 1900 CAGACAGTGAATCTGCTGCTGCAAGCTGATGTAAATTAAGTTCTAGA 1951
Db 421 CAGACAGTGAATCTGCTGCTGCAAGCTGATGTAAATTAAGTTCTAGA 472

RESULT 11
US-10-027-632-302414
Sequence 302414, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632


```

RESULT 13
US-10-027-632-302414
; Sequence 302414, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping
; TITLE OF INVENTION: Polymorphisms in the
; FILE REFERENCE: 108827.139
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218

```



```

; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 302414
; LENGTH: 472
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-302414

Query Match
Best Local Similarity 18.4%; Score 471.2; DB 14; Length 472;
Matches 470; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1480 TGCTCTCTCCGTCGAGAAATCAAGCAGGAAACAGACGACTGAGCCTTCTGCTCCAG 1539
Db 1 TGCTCTCTCCGTCGAGAAATCAAGCAGGAAACAGACGACTGAGCCTTCTGCTCCAG 60

QY 1540 CTGGCTCTTAGCTGGAAGCCAGCCAGCGTTTCTTACCAACACACCATGCCCTCCACC 1599
Db 61 CTGGCTCTTAGCTGGAAGCCAGCCAGCGTTTCTTACCAACACACCATGCCCTCCACC 120

QY 1600 TGACTTTGGCTGGAGACTGATCCTCTCTGTTAAATCTGCGCGGTGCTGGAAGCTG 1659
Db 121 TGACTTTGGCTGGAGACTGATCCTCTCTGTTAAATCTGCGCGGTGCTGGAAGCTG 180

QY 1660 GACGFTGGAGGACCTGCTGGGTCTCTGGGACCCGCTGTTGCTTCTGCGCTCCCTGTT 1719
Db 181 GACGFTGGAGGACCTGCTGGGTCTCTGGGACCCGCTGTTGCTTCTGCGCTCCCTGTT 240

QY 1720 GGAAGGCTATATGACGGGCGCCTGAGCCGCCAGAACTGCTGTGTAACACCTTTTC 1779
Db 241 GGAAGGCTATATGACGGGCGCCTGAGCCGCCAGAACTGCTGTGTAACACCTTTTC 300

QY 1780 CAGCCAGAGTCCCAAGCTGGAAGCTGAGCTGCTGCTTCTTCTTAAAGTGGCTCCCC 1839
Db 301 CAGCCAGAGTCCCAAGCTGGAAGCTGAGCTGCTGCTTCTTCTTAAAGTGGCTCCCC 360

QY 1840 CCGACCGCCAGCGCCCTCAGTTGCGAGGATGGGGCCACCACTGTCTACATGTGGAATA 1899
Db 361 CCGACCGCCAGCGCCCTCAGTTGCGAGGATGGGGCCACCACTGTCTACATGTGGAATA 420

QY 1900 CAAGCAGTGAATCTGCTGCTGCAACGAGTCATGTAATTAAGTTCTAGA 1951
Db 421 CAAGCAGTGAATCTGCTGCTGCAACGAGTCATGTAATTAAGTTCTAGA 472

```

```

RESULT 14
US-09-918-995-22234
; Sequence 22234, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22234
; LENGTH: 492
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature

```

```

; LOCATION: (1)...(492)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-22234

Query Match
Best Local Similarity 17.1%; Score 438; DB 11; Length 492;
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1947 CTAGCAGAGCTCTCTGAGCAGGATAAGTCCCTCCCTGACAGTGTGTGTGGTGGGGCAG 2006
Db 55 CTAGCAGAGCTCTCTGAGCAGGATAAGTCCCTCCCTGACAGTGTGTGTGGTGGGGCAG 114

QY 2007 CCTCTGCTCAAAAATTCACCAAGCAGAGTCTCTCAGCCTCATGTGTGGTCTCTGTC 2066
Db 115 CCTCTGCTCAAAAATTCACCAAGCAGAGTCTCTCAGCCTCATGTGTGGTCTCTGTC 174

QY 2067 TCTCTCTAGCTCCCGAGGATGTTGGGACCCAGCTTGTCTCGGACGCTAAGAGCAGTG 2126
Db 175 TCTCTCTAGCTCCCGAGGATGTTGGGACCCAGCTTGTCTCGGACGCTAAGAGCAGTG 234

QY 2127 ACCAGGATGTGATTTTGGCGACCTGTGTGTGGTGGCTTCTGAGTGTCTTGTGTGTA 2186
Db 235 ACCAGGATGTGATTTTGGCGACCTGTGTGTGGTGGCTTCTGAGTGTCTTGTGTGTA 294

QY 2187 GGAAGTGTGATTTTGGCGACCTGTGTGTGGTGGCTTCTGAGTGTCTTGTGTGTA 2246
Db 295 GGAAGTGTGATTTTGGCGACCTGTGTGTGGTGGCTTCTGAGTGTCTTGTGTGTA 354

QY 2247 CTGAGTGTGACATTTGGCTCTGCTAGCTGTCTCTGAAGCCCGCTAAATAATTCATCC 2306
Db 355 CTGAGTGTGACATTTGGCTCTGCTAGCTGTCTCTGAAGCCCGCTAAATAATTCATCC 414

QY 2307 AAGATTCCTTTAGTAAAGGTCCAGTCTGAGTGGAGCCCTTAGAGAGCTGGGCTTG 2366
Db 415 AAGATTCCTTTAGTAAAGGTCCAGTCTGAGTGGAGCCCTTAGAGAGCTGGGCTTG 474

QY 2367 TATGTTCTTTTGGCCCTT 2384
Db 475 TATGTTCTTTTGGCCCTT 492

```

```

RESULT 15
US-09-789-919-55
; Sequence 55, Application US/09789919
; Patent No. US20020064855A1
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ihor
; APPLICANT: Moore, Kateri
; TITLE OF INVENTION: GENES THAT REGULATE HEMATOPIETIC BLOOD FORMING STEM
; FILE OF INVENTION: CELLS AND USES THEREOF
; FILE REFERENCE: 2275-1-005
; CURRENT APPLICATION NUMBER: US/09/789,919
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 55
; LENGTH: 3590
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-789-919-55

Query Match
Best Local Similarity 16.2%; Score 413.8; DB 9; Length 3590;
Matches 735; Conservative 0; Mismatches 462; Indels 24; Gaps 2;

QY 1 AGCAGCTCCCACTCTATGACAGTGGGAGGATGTGATCAAGGGGATGAAGTGGAGTGC 60
Db 750 AGCAGCTCCCACTCTATGACAGTGGGAGGATGTGATCAAGGGGATGAAGTGGAGTGC 809

QY 61 TCAACAGTGTGCTGTCTCCAGCCGGGTGTACTGGATCGCTCTCTCATCCAGACAG 120
Db 810 CCAATACAGACTGCAGTCTACTCTACCAAGGCTTCTGTGATTGCTGGAATTATAAATTAG 869

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 4, 2004, 13:33:09 ; Search time 5218 Seconds
(without alignments)
11900.716 Million cell updates/sec

Title: US-10-031-915-90

Perfect score: 2555

Sequence: 1 agcagctccactctatgac.....ttgcttgagaaaaa 2555

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pin:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_nam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rtd:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1203.6	47.1	3441	11	BC023933
2	1200.4	47.0	3422	11	AK029115
3	1200.4	47.0	3431	11	AK036510
4	960.4	37.6	1201	13	BX446508

C	5	932.8	36.5	1065	13	BX360491
	6	878	34.4	1201	9	AL560108
	7	872	34.1	1080	12	BM909451
	8	794.6	31.1	900	13	BU184978
	9	783	30.6	927	13	BQ927459
	10	773	30.3	893	14	CA454956
	11	743.2	29.1	876	13	BQ217463
	12	739.8	29.0	851	12	BI092956
	13	737.6	28.9	1057	10	BG396658
	14	734.2	28.7	914	10	BG400812
	15	733.4	28.7	855	12	BI919105
	16	720.8	28.2	915	13	BX445939
	17	704.6	27.6	1201	9	AL581617
C	18	692.6	27.2	736	13	BU634412
C	19	692.6	27.1	739	14	CD365541
C	20	691.4	27.1	821	13	BQ573608
C	21	690.6	27.0	733	10	BG685391
C	22	689.8	27.0	721	12	BM666677
C	23	688.4	26.9	695	12	BM666631
C	24	684	26.8	684	14	CB154753
C	25	682.8	26.7	703	14	CA446367
C	26	681.4	26.7	875	13	BX327327
C	27	673.8	26.4	931	10	BG757068
C	28	673.6	26.4	686	12	BM681205
C	29	670.8	26.3	691	13	BU742966
C	30	662	25.9	1089	12	BM479405
C	31	659	25.8	902	13	BQ438013
C	32	657.2	25.7	667	13	BU737070
C	33	643.8	25.2	664	14	CD370874
C	34	641	25.1	900	10	BE729647
C	35	628.8	24.6	762	13	BQ572663
C	36	614.6	24.1	688	10	BE789481
C	37	613.8	24.0	622	10	BF390805
C	38	607	23.8	749	10	BE386405
C	39	604	23.6	605	12	BM726551
C	40	593	23.2	859	14	CA750138
C	41	588	23.0	854	10	BG258301
C	42	575	22.5	693	10	BE280391
C	43	574.6	22.5	2032	11	BC023676
C	44	571.4	22.4	784	14	CA510759
C	45	569.4	22.3	599	12	BM973511

ALIGNMENTS

BC023933 3441 bp mRNA linear HTC 23-SEP-2002
Mus musculus, Similar to hypothetical protein MGC31247, clone
IMAGE:5325131, mRNA.
ACCESSION BC023933
VERSION BC023933.1 GI:23271699
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3441)
Strausberg,R.
Direct Submission
Submitted (05-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),

REMARK
COMMENT

Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mc@nih.gov
Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Green, X., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lario, P., Legaspi, R.,
Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, K., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 56 Row: 0 Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis

This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers

1. 3441

/organism="Mus musculus"

/mol_type="mRNA"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:5325131"

/cissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."

/clone_lib="NCI CGAP_Mam6"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

BASE COUNT 817 a 922 c 970 g 732 t

ORIGIN

Query Match 47.1%; Score 1203.6; DB 11; Length 3441;
Best Local Similarity 86.5%; Pred. No. 62-273;
Matches 1353; Conservative 0; Mismatches 204; Indels 7; Gaps 2;

1 AGCACGTCCTCTATGACAGTGGAGGATGATCAAGGGATGAAGTGGAGTGC 60

649 AACATGTACCTCTATGACAGTGGAGAGCTCATGAGGGATGAAGTGGAGTGC 708

61 TCAACAGTATCTGTGCTCCCGAGCGGTGTACTGATGCTCTCTCATCCAGACAG 120

709 TCAACAGAGATCTGTGCTCCCGAGCGGTGTACTGATGCTCTCTCATCCAGACAG 768

121 CAGGATATCGGCTGTCTCGTATGAAGGCTTTGAAAATGACCCAGCCATGACTTCT 180

769 CTGGTACCGGTGTCTCGATATGAAGCTTTGAAAATGACCCAGCTCATGACTTCT 828

181 GGTGCAACCTGGGAACAGTGGATGTCACCCCATTTGGTGTGTGTCATCAACAGCAAGA 240

829 GGTGCAACCTGGGAACAGTGGATGTCACCCCATTTGGTGTGTGTCATCAACAGCAAGA 888

241 TCCTAGTGTCCCGAGGACCTCATGTCACAGTTCACCGATGGAGGCTTACTCTATGA 300

889 TCCTAGTGTCCCGAGGACCTCATGTCACAGTTCACCGATGGAGGCTTACTCTATGA 948

301 AACCGTGTGGTGTCCAGGACCTTCCCGTGGATTTCCACATCAAGATGGTGGAGAGCA 360

949 AGCGGTGTGGTGTCCAGGACCTTCCCGTGGATTTCCACATCAAGATGGTGGAGAGCA 1008

361 TGAAGTACCTTTAGGACAGGACCTTCCCGTGGATTTCCACATCAAGATGGTGGAGAGCA 420

1009 TGAAGTACCTTTCCAGAGGACCTTCCCGTGGATTTCCACATCAAGATGGTGGAGAGCA 1068

421 GCATCTCGATGCTGTGGTGGACACAGTAATCGGGGTTCGCTTACGGCTCTCTTACGAGG 480

1069 GTACCCGATGCTGTGGTGGACACAGTAATCGGGGTTCGCTTACGGCTCTCTTATGAGG 1128

481 ATGGTGAAGTGAACACAGTCTTGGTGGACACAGTAATCGGGGTTCGCTTACGGCTCTCTTACGAGG 540

1129 ATGGTGAAGTGAACACAGTCTTGGTGGACACAGTAATCGGGGTTCGCTTACGGCTCTCTTACGAGG 1188

QY 541 GTTGGTCAACGAGTGTGGCCACGGCATCAAGATGTCAGAGAGCGGAGTGAATGCGCC 600

Db 1189 GTTGGTCCCGCGTGTGGCCACGGCATCAAGATGTCAGAGAGCGGAGTGAATGCGTC 1248

QY 601 ATCAACCCACCTTCCGGAAGATCTACTGTGATCCGCTTCTTACCTCTTCAAGAGGTAC 660

Db 1249 ATCAACCCACCTTCCGGAAGATCTACTGTGATCCGCTTCTTACCTCTTCAAGAGGTAC 1308

QY 661 GAGCAGTCTACACAGAGGCGGTGTTTTCAGAGAGGATGAAGTGGAGGCGCATTCAC 720

Db 1309 GCGCTGCTTACACAGAGGCGGTGTTTTCAGAGAGGATGAAGTGGAGGCGCATTCAC 1368

QY 721 CCTTGAATCTGGGCAACATCTGGTGGCAACTCTGTGTAAAGTTCCTCTGGATGGATACC 780

Db 1369 CTCTGAATCTGGGCAAGTATCTGTGTAGCAACCATCTGCAAGGTGCTCTTGGATGGTACC 1428

QY 781 TGATGATCTGTGTGAGCGGGGGCCCTCCACAGATGGCTTGGACTGGTGTCTGATCATG 840

Db 1429 TGATGATCTGTGTGAGCGGGGGCCCTCCACAGATGGCTTGGACTGGTGTCTGATCATG 1488

QY 841 CCTCTTCCCAAGCCATCTTCCCGGCCACCTTCTGTGAGAGATGACATTTGAGCTCACAC 900

Db 1489 CCTCTTCCCAAGCCATCTTCCCGGCCACCTTCTGTGAGAGATGACATTTGAGCTCACAC 1548

QY 901 CGCCAAAGGTTATGAGGACACAGCTTTTAACTGGGAGAACTACTTTGGAGAGACCAAGT 960

Db 1549 CCCCAGAGGTTATGAGACACAGCTTTTCCCTGGGAGACCTACTTTAGAGAGACCAAGT 1608

QY 961 CGAAGCGCTCCATCGAGACTCTTTAACTGATGATTTGCCCAACCATGGCTTCAAGGTGG 1020

Db 1609 CAAAGCTCTCCAGCAAGACTTTTAACTGATGATTTGCCCAACCATGGCTTCAAGGTGG 1668

QY 1021 GCATGAAGTGGAGCGCTGGAGCTTGAAGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1080

Db 1669 GCATGAAGTGGAGCGCTGGAGCTTGAAGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1728

QY 1081 AACGAGTGTGTCATCGGCTCTCAGATCCACTTTTGAACGCTGGGACAGCGAGTACGAC 1140

Db 1729 AGCGGTGTGTCATCGGCTCTCAGATCCACTTTGACGCTGGGACAGCGAGTACGAC 1788

QY 1141 AGTGGTGAAGTCCGAGTCCCAGACATCTACCCGCTCGGCTGGTGTGAGCTCACCGGT 1200

Db 1789 AGTGGTGAAGTCCGAGTCCCAGACATCTACCCGCTCGGCTGGTGTGAGCTCACCGGT 1848

QY 1201 ACCAGTCCAGCTCTGTGTGGGAGAACCGGACACACCGCTGAAGGCCAAGAGGCA 1260

Db 1849 ACCAGTCCAGCTCTGTGTGGGAGAACCGGACACACCGCTGAAGGCCAAGAGGCA 1908

QY 1261 CAAAGAGAGAGAGAGAGAGTGTGGAGAGAGAGAGAGAGTGTGGAGAGAGAGAGAG 1320

Db 1909 CAAAGAGAGAGAGAGAGAGTGTGGAGAGAGAGAGAGAGTGTGGAGAGAGAGAGAG 1968

QY 1321 GACCCCTCAGACAGGCTCCAGAGAGCCCTTCTGAGAGAGAGAGAGAGAGAGAGAG 1380

Db 1969 GACCCCTCAGACAGGCTCCAGAGAGCCCTTCTGAGAGAGAGAGAGAGAGAGAGAG 2027

QY 1381 AGATCTCGTGGAGAGAGTGTCTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440

Db 2028 -----GTCTCGAGAGAGAGTGTCTGTGGAGAGAGAGAGAGAGAGAGAGAGAG 2082

QY 1441 TAGAGTGGCTTCGCGGAGAGAGAGTTCAGAGTCCAGAGAGAGAGAGAGAGAGAGAG 1500

Db 2083 AGGAGATTTCTCGGCTGACAGAGTGGCCAGTCCACAGCTGCTCTTCCATTTAGAGAG 2142

QY 1501 TCAAGCAGAGAGAGAGAGAGTTCAGAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1559

Db 2143 TCAGCAGAGAGAGAGAGAGTTCAGAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 2202

QY 1560 CAGC 1563

Db 2203 AAGC 2206

RESULT 2

AK029115

LOCUS

DEFINITION

AK029115 3422 bp mRNA linear HTC 05-DEC-2002
Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
library, clone:4732493N06 product:H-L(3)MBT-LIKE PROTEIN
(HYPOTHETICAL 79.1 KDA PROTEIN) homolog [Homo sapiens], full insert
sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AK029115.1 GI:26325101

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hitamoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imorani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kurahara, C., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kawai, J., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohata, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M., and Hayashizaki, Y.

TITLE

JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-rse@sc.riken.go.jp,
URL: http://genome-gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues

Please visit our web site for further details.

URL: http://genome-gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/

FEATURES

source

1..3422
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:4732493N06"
/db_xref="taxon:10090"
/clone="4732493N06"
/tissue_type="skin"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days neonate"
58..2169
/note="unnamed protein product; H-L(3)MBT-LIKE PROTEIN
(HYPOTHETICAL 79.1 KDA PROTEIN) homolog [Homo sapiens]
(SPT969R5, evidence: FASTA, 90.9%ID, 100%length,
match=2109)
putative"
/codon_start=1
/protein_id="BAC26305.1"
/db_xref="GI:26325102"
/translation="MEKPRGTPEAPSEPEMEEDDLDFGYSFRSYNSAGSES
SSYLEESSEAEENEDREAGELPTSPHLFSSANNRSLDGSSEPAVCEMCGIVGTREAF
PSKTRFCVSCRSYSSNSKASILARLQGGPKTKAKVLHKAWSAKIGFLHAQG
TGGLADGTGTGQDALVGLGDFWCKFLKDSYKAPVSCFQGVPLVDQWEDVMGKVEV
LNSDAVLPSRVNIAIVIAQGYRVLRYEGENDASHDFWNLGTVDPHPIGWCAIN
SKLVPRTHAKFTDWKSLMKRLVGSRTLPADPHIKMVSMEKYPFRQGMLEVVDDK
TQVSRMAVVDITVIGRLLYEDGSDDDFCFHWMSPLIHPVGSRRVGHGIVKSD
RCDMSHPHTFRKICADVPLFKVRAVYTEGGWFEGEKMLEIDLPLGLSICVATI
CKVLLDGYSLICVDGSPSTDGFWFCFTHASSHAIFFATPCQKNDILPTPKGYETQPF
AVETYLETKSKAAPLPMDCPNHGVKMKLEAVDLMEPLLICVATVKVVRHLL
SHFDGMDNEYDQWVDCSPDIPVYGCWELTCVQLQPPVSAEPNTPOKQDTPKKKK
QKQKKEKLPSPAKTRPLRGSKPILLEDLNLGLVSEPVDPDIIAIVCKEERQDSSL
DASPSQLPLPESINQERN"

CDS

polyA_signal
/note="putative"
3404..3409
polyA_site
/note="putative"
3422

BASE COUNT 800 a 917 c 966 g 739 t

ORIGIN

Query Match 47.0%; Score 1200.4; DB 11; Length 3422;
Best Local Similarity 86.4%; Pred. No. 3.4e-272;
Matches 1351; Conservative 0; Mismatches 206; Indels 7; Gaps 2;
QY 1 AGCACGTCCCACTTATGACGAGGATGTGATGAAGGATGAGGTGGAGGTGC 60

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Db 653 AACATGTACCCCTCTATGACCAAGTGGGAAGACGTCATGAAGGGATGAAGTGGAGTGC 712
Qy 61 TCAACAGTGTATGCTGCTCCAGCGGGTGTACTGATCCCTCTGTATCCTCAGACAG 120
Db 713 TCAACAGCATGTGTGCTCCAGCGGGTGTACTGATCCCTCTGTATCCTCAGACAG 772
Qy 121 CAGGGTATCGGGTGTGCTTCGGTATGAAGGGTGTGAAATGACGCCAGCCATGACTTCT 180
Db 773 CTGGGTACCGGGTGTGCTCCGATATGAAGGCTTTGAAATGACGCCAGTCACTTCT 832
Qy 181 GGTGCAACCTGGAAACAGTGGATGTCCACCCATCTGGCTGGTGGCTGCAATCAACAGCAGA 240
Db 833 GGTGCAACCTGGAAACAGTGGATGTCCACCCATCTGGCTGGTGGCTGCAATCAACAGCAGA 892
Qy 241 TCTAGTGTCCCGCCACGAGCATCCATCCAGTTCACCGACTGGAAGGCTACCTCATGA 300
Db 893 TCTGTGTACTCCACGACCATCCATCCAGTTCACCGACTGGAAGGCTACCTCATGA 952
Qy 301 AACGGCTGTGGCTCCAGGACGCTTCCCGTGGATTTCCACATCAAGATGTTGGAGACA 360
Db 953 AGCGGTGTGGTTCAGGACACTTCTCTGCAGACTTCCATATCAAGATGTTGGAGACA 1012
Qy 361 TGAAGTACCCCTTTAGCGAGGCGATGCGCTGGAAGTGTGGACAAAGTCCAGGTGTAC 420
Db 1013 TGAAGTACCCCTTTAGCGAGGCGATGCGCTGGAAGTGTGGACAAAGTCCAGGTGTAC 1072
Qy 421 GCACTCGCATGCTGTGGTGGACACAGTAAATCGGGGTGCGCTACGGTCTCTACGAGG 480
Db 1073 GTACCGCATGCGCGTGTGGACACAGTAAATCGGGGTGCGCTACGGTCTCTACGAGG 1132
Qy 481 ATGGTGACAGTGAACGACACTTCTGGTGCCACATGTGGAGCCCTGTATCCACCGAGTG 540
Db 1133 ATGGTGACAGTGAACGACACTTCTGGTGCCATGTGGAGTCCCTGTATCCACCGAGTG 1192
Qy 541 GTTGTGTACAGCTGTGGGCCCGCATCAAGATGTCAAGAGGCGAAGTGCATGGGCC 600
Db 1193 GTTGTGTCCCGCGTGTGGGCCCGCATCAAGATGTCAAGAGGCGAAGTGCATGGTCT 1252
Qy 601 ATCACCCACCTTCGGAGATCTACTGTAGTGGCTTCTTACTCTTCAAGAGGTAC 660
Db 1253 ATCACCCACCTTCGGAGATCTACTGTAGTGGCTTCTTACTCTTCAAGAGGTAC 1312
Qy 661 GAGCAGTCTACACAGAGCGGTGTGGTTGAGAGAGGATGAAGTGGAGGCCATTGACC 720
Db 1313 GCGCTGTCTACAGAGAGTGTGGTTCGAGAGAGGATGAAGTGGAGGCCATTGACC 1372
Qy 721 CCTGAATCTGGCAACATCTCGTGGCAACTGTCTGTAAAGTCTCTCTGGATGGATACC 780
Db 1373 CTCTGAATCTGGCAAGTATCTGTAGCAACCATCTGCAAGGTGCTCTTGGATGGTTACC 1432
Qy 781 TGATGATCTGTGGACGGGGGCCCTCCACAGATGGCTTGGACTGTTCTGTACCATG 840
Db 1433 TGATGATCTGTGGATGGGGGCCCTCCACGATGGCTCCGATGGTCTGTCTACCAAG 1492
Qy 841 CTCTTCCACGCCCATCTTCCCGGCCACCTTCTGTCAAGAAATGACATGAGCTCAC 900
Db 1493 CTCTTCCCATGCCATCTTCCAGCCACCTTCTGCCAAAAGATGACATGAGCTCAC 1552
Qy 901 CGCCAAAAGTTATGAGGACAGATTTCACTGGAGAACTACTTGGAGAAAGCAAGT 960
Db 1553 CCCCAAAAGGTATGAGACACAGCTTTTGGCTGGAGACCTTACTTAGAAGAACCAAGT 1612
Qy 961 CGAAAGCCCGCTCCATCGAGACTCTTTTAAATGATGATGGTCCCAACCATGGCTTCAAGTGG 1020
Db 1613 CAAAAGCTCTCCAGCAAGACTCTTTTAAATGATGATGGTCCCAACCATGGCTTCAAGTGG 1672
Qy 1021 GCATGAAGCTGAGCGCGTGGACCTGTATGGAGCCCGCGCTCATCTGTGTGGCCACGGTGA 1080
Db 1673 GCATGAAGCTGAGCGCTGGACCTGTATGGAGCCCGCGCTCATCTGTGTGGCCACCTGTGA 1732
Qy 1081 AACGAGTGTGTATCGGCTCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACGACC 1140

Db 1733 AGCGGTGTGTATCGGCTCTCAGCATCACTTTCAGCGCTGGGACAATGATGATGATC 1792
Qy 1141 AGTGGTGGACTCGAGTCCCGAGACATCTACCCCTCGCTGGTGTGAGTCTACCGGCT 1200
Db 1793 AGTGGTGGACTCGAATCCCGGACATCTACCTGTGCTGGTGTGAGTCTACCGGCT 1852
Qy 1201 ACCAGTCTCCAGCTCTCTGTGGCCGAGAAACCGGCCACACCGCTGGAAGGCCAAAGAGGCCA 1260
Db 1853 ACCAGTCTCCAGCTCTCTGTGGCCGAGAAACCGGCCACACCTCAGAAAGGCCAAGCACCA 1912
Qy 1261 CAAAGAAAGAAAGAAACAGTTTGGGAGAAAGAAAGAAATCCGCCCACTAAGACGC 1320
Db 1913 CAAAGAAAGAAAGAAACAGTTTGGGAGAAAGAAAGAAATCCCATCAGCCAAAGACTC 1972
Qy 1321 GACCCCTCAGACAGGGGTCCAGAAAGCCCTGTGGAGGACGACCCCTCAGGGTGCAGGA 1380
Db 1973 GCGCCCTCAGACAGGGCTCCAGAAACCCCTTACTGGAGGACCACTTGGGCTTTGGGG- 2031
Qy 1381 AGATCTCTGGAGGCTTCTCTGGGAGATCATCTGTGGTGTGAAGGAAGAGCATC 1440
Db 2032 -----GTCTCGGAACAGTTCTCTGATGACATTATCTGTGTGTGAAGGAGGAGCACC 2086
Qy 1441 TAGAGTGGCTCGCCGCAAGGCTTCAAGTCCAGAGCTGCTGTCTCGTCGGAACA 1500
Db 2087 AGGACATTTCTCTGTGAGAGTCCCGAGTCCAGAGTCTTCCATTTGAGAGCA 2146
Qy 1501 TCAAGCAGGAAACAGACGACTGAGCCCTT-CCTGCTCCAGCCCTGGCTTCTAGCTGGAAGC 1559
Db 2147 TCAAGCAGGAGAGGAACAATGAGACTTCCCTGGCATCAGCTTGGACCCCTTACTGAAGCC 2206
Qy 1560 CAGC 1563
Db 2207 AAGC 2210

RESULT 3
AK036510
LOCUS

DEFINITION

AK036510 3431 bp mRNA linear HTC 05-DEC-2002
Mus musculus adult male bone cDNA, RIKEN full-length enriched
library, clone:19830123N10 product:H-L(3)MBT-LIKE PROTEIN
(HYPOTHEICAL 79.1 KDA PROTEIN) homolog (Homo sapiens), full insert
sequence.

ACCESSION

AK036510

VERSION

AK036510.1 GI:263331451

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

REFERENCE

1 Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999).

2

10349636

REFERENCE

AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalisation and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

3

11042159

REFERENCE

AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,

Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, S., Watahiki, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer

```

1. 3431
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:9830123N10"
/db_xref="taxon:10090"
/clone="9830123N10"
/sex="male"
/tissue_type="bone"
/clone_lib="RIKEN full-length enriched mouse cDNA library"

```

QY	601	ATCACCCACCTTCGGAAGATCTACTGTGATCGGTCCTTACTCTTCAAGAGGTAC	660
DB	1261	ATCACCCACCTTCGGAATACTGTGATGTGTACTTACTCTTCAAGAAGTCC	1320
QY	661	GAGCAGTCTACACAGAAGCGGTGGTTTGAGGAAGGATGAAGCTGAGGCCATTGACC	720
DB	1321	CGCTGTCTACACAGAAGTGGCTGGTTTCGAGGAAGGAATCAAACTAGAGCCATTGACC	1380
QY	721	CCCTGAATCTGGGCAACATCTGCGTGGCAACTGTCTGAAGGTTCTCTCGATGGAATAC	780
DB	1381	CTCTGAATCTGGGCAGTATCTGTGTAGCAACCACTCTGCAAGGTGCTCTTGGATGTTTACC	1440
QY	781	TGATGATCTGTGTGGACGGGGGCCCTCCACAGATGCTTGGACTGTTCTGCTACCACATG	840
DB	1441	TGATGATCTGTGTGGATGGGGGGCCCTCCACGATGGCTCCGATGGTTGTTCTGCTACCACG	1500
QY	841	CCTCTTCCACGCCATCTTCCGGCCACCTTCTGTGAGAAGATGACATTGAGCTCACAC	900
DB	1501	CCTCTCCCATGCCATCTTCCAGCACCTTCTGCCAAGAAGATGACATTTGAGCTCACAC	1560
QY	901	CGCCAAAGGTTATGAGGCACAGACTTTCACTGGGAGAACTACTTGGAGAAGCAACAAGT	960
DB	1561	CCCCAAAGGGTATGAGACACAGCCTTTTGCCTGGGAGACTACTTAGAGAAGACCAAGT	1620
QY	961	CGAAGCGGCTCCATCCAGACTCTTTAAATGATTCGCCAAACCACCTGGCTTCAAGGTGG	1020
DB	1621	CAAAAGCTGCTCCAGCAAGACTCTTTAAATGAGACTGCCCAACCACTGGCTTCAAGGTGG	1680
QY	1021	GCATGAGCTGGAGCCGCTGGACCTGATGAGCCCGGCTCATCTGTGTGGCCACGGTGA	1080
DB	1681	GCATGAGCTGGAGGCTGTGGACTGATGAGCCCGGCTCATCTGTGTGGCCACTGTGA	1740
QY	1081	AACGAGTGTGCATCGGCTCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACGACC	1140
DB	1741	AGCGGTGGTGCATCGSCTCCTCAGCATCCACTTCGACGGCTGGGACATGATATGATC	1800
QY	1141	AGTGGTGGACTGGAGTCCCGACACATCTACCCGCTCGSCTGGTGTGAGCTCACCGGT	1200
DB	1801	AGTGGTGGACTCGGAATCCCGGACATCTACCCCTGTGCGTGGTGTGAGCTCACCGGT	1860
QY	1201	ACCAGTCTCAGSCTCTCTGTGGCGCGAGAACCGGCCACACCGCTGGAAGGCCAAAGAGCCA	1260
DB	1861	ACCAGTCTCAGCGCCTGTCTCCGAGAACCAACACACCTCAGAAAGCCAGGACACCA	1920
QY	1261	CAAGAGAAAGAAAGAACAGTTTGGAGAAAGAAAGAAATCCGCCCATTAAGACGC	1320
DB	1921	CAAGAGAAAGAAAGAACAGTTTGGAGAAAGAAAGAAATCCCATCAGCCAAAGACTC	1980
QY	1321	GACCCCTCAGACAGGGGTCCAAAGACCCCTGTGGAGGACGACCCCTCAGGGTGCACGA	1380
DB	1981	GGCCCTCAGACAGGGCTCCAAAGAACCCCTTACTGGAGGACACCTTGAGGCTTTGGGG-	2039
QY	1381	AGATCTCGTCGAGCTGTTCCTGGCGAGATCATCTGTGTGGTGTGAAGGAAGACATC	1440
DB	2040	-----GTCTCGGAACAGTTCTCTGTGACATATCGCTGTGTGTGAAGGAGGACACC	2094
QY	1441	TAGACGTGGCCTCGCCGACAAAGGCTTCAAGTCCAGAGTGCCTGTCTCCGTCGAAACA	1500
DB	2095	AGGACATTTCTCGCTTGACAGGTGCCCCAGTCCACAGCTGCCTCTTCCATTGAGACA	2154
QY	1501	TCAAGCAGGAACACAGACGACTGAGCCTT--CCTGCTCCAGCCTGGCTTCTAGCTGAAGC	1559
DB	2155	TCAAGCAGGAGGAACAACCTGAGACTTCCTGGCATCAGCCTGGACCCCTAACTGAAGCC	2214
QY	1560	CAGC	1563
DB	2215	AAGC	2218

RESULT 4	1201 bp	mRNA	linear	EST 22-MAY-2003
BX446508				
LOCUS	BX446508			

DEFINITION	EX44508 Homo sapiens PLACENTA Homo sapiens cDNA clone CLOBA010ZA01
ACCESSION	S-PRIME, mRNA sequence.
VERSION	EX44508
KEYWORDS	EST.
SOURCE	EX44508.1 GI:31031730
ORGANISM	Homo sapiens (human)
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1. (bases 1 to 1201)
TITLE	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
JOURNAL	Full-length cDNA libraries and normalization
COMMENT	Unpublished Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 10409.f For more information about this cluster, see http://www.genoscope.cns.fr/ cgi-bin/cluster.cgi?seq=CLOBA010ZA01R1&cluster=10409.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Paradise Avenue Genoscope sequence ID : CLOBA010ZA01RP1. Location/Qualifiers 1. .1201 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CLOBA010ZA01" /tissue_type="PLACENTA" /clone_lib="Homo sapiens PLACENTA" /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, into double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."
BASE COUNT	260 a 340 c 315 g 247 t 39 others
ORIGIN	

Query Match	37.6%;	Score	960.4;	DB	13;	Length	1201;
Best Local Similarity	96.9%;	Pred. No.	1.3e-215;				
Matches 1034;	Conservative	16;	Mismatches	10;	Indels	7;	Gaps
QY	1091	GCATCGGCTCCTCAGCATCCACTTTGACGGCTGGGACGCGAGTACGACCACTGGGTGGA	1150				
Db	29	GGATCGGCTCCTCAGCATCCACTTTGACGGCTGGGACGCGAGTACGACCACTGGGTGGA	88				
QY	1151	CTGCGAGTCCCAGAGACATCTACCCCGTGGCTGGTGTGAGCTTCACCGGTACGAGTCCA	1210				
Db	89	CTGCGAGTCCCAGAGACATCTACCCCGTGGCTGGTGTGAGCTTCACCGGTACGAGTCCA	148				
QY	1211	GGCTCTGTGGCGGAGAACCGGCCACACCCGTGTGAAGGCCAAGAGGCCACAAAGAGAA	1270				
Db	149	GGCTCTGTGGCGGAGAACCGGCCACACCCGTGTGAAGGCCAARAGGCCACAAAGAGAA	208				
QY	1271	AAAGAAAACAGTTTGGGAAGAAAGAAAAGAAATCCCGGCCACTAAGACGCGACCCCTCAG	1330				
Db	209	AAAGAAAACAGTTTGGGAAGAAAGAAAAGAAATCCCGGCCACTAAGACGCGACCCCTCAG	268				
QY	1331	ACAGGGGTCCAAGAACCCCTGTCTGGAGGAACGACCCCTCAGGGTGCCAGGAAGATCTCGTC	1390				
Db	269	ACAGGGGTCCAAGAACCCCTGTCTGGAGGACGACCCCTCAGGGTGCCAGGAAGATCTCGTC	328				
QY	1391	GGAGCCTGTTCCTGGCGGAGATCATTTGCTGTGGTGTGAAGGAAGAGCATCTAGACGTGGC	1450				
Db	329	GGAGCCTGTTCCTGGCGGAGATCATTTGCTGTGGTGTGAAGGAAGAGCATCTAGACGTGGC	388				
QY	1451	CTCGCCCGACAAAGGCTTCAAGTCCAGAGCTGCTGTCTCCGTTCGAGAACATCAAGACAGA	1510				
Db	389	CTCGCCCGACAAAGGCTTCAAGTCCAGAGCTGCTGTCTCCGTTCGAGAACATCAAGACAGA	448				


```

QY 1511 AACAGCAGTGGCTTCTGCTCAGCCTGGCTTAGCTGGAAGCCAGCCAGCCT 1570
Db 449 AACAGCAGTGGCTTCTGCTCAGCCTGGCTTAGCTGGAAGCCAGCCAGCCT 508
QY 1571 TTCTCTACCAACCAACCACTCCACCTGACTTTGGCTTGAGACTGATCTCTCTG 1630
Db 509 TTCTCTACCAACCAACCACTCCACCTGACTTTGGCTTGAGACTGATCTCTCTG 568
QY 1631 GTAATTTGCGCGTGTGTGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTG 1690
Db 569 GTAATTTGCGCGTGTGTGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTG 628
QY 1691 ACCCGCTGTGCTTCTGCTCCCTGCTGGAAGCTGGAAGCTGGAAGCTGGAAG 1750
Db 629 ACCCGCTGTGCTTCTGCTCCCTGCTGGAAGCTGGAAGCTGGAAGCTGGAAG 688
QY 1751 CCCAGAACTCGTGTGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGA 1810
Db 689 CCCAGAACTCGTGTGAAGCTGGAAGCTGGAAGCTGGAAGCTGGAAGCTGGA 748
QY 1811 TGCGTCTCTCTTGAAGTGGCTTCCCGCAGCCGCGCAGCCGCTGAGTTGCCAGGA 1870
Db 749 TGCGTCTCTCTTGAAGTGGCTTCCCGCAGCCGCGCAGCCGCTGAGTTGCCAGGA 808
QY 1871 TGCGGCGCAGCTGTGCACTGTGAAGTGAAGCTGAAGCTGAAGCTGAAGCTGA 1930
Db 809 TGCGGCGCAGCTGTGCACTGTGAAGTGAAGCTGAAGCTGAAGCTGAAGCTGA 868
QY 1931 TCAATGAATTAAGTTCTAGAGCAGCTCTGAGCAGGATAGGTCCTGACAGTGA 1990
Db 869 CCAATGAATTAAGTTCTAGAGCAGCTCTGAGCAGGATAGGTCCTGACAGTGA 928
QY 1991 TGTGTGCTGGGCGGCGCTCTGCTCAGGATGTTGGGACCCAGCTTGTCTCGG 2050
Db 929 TGTGTGCTGGGCGGCGCTCTGCTCAGGATGTTGGGACCCAGCTTGTCTCGG 986
QY 2051 TGTGTGCTGGGCGGCGCTCTGCTCAGGATGTTGGGACCCAGCTTGTCTCGG 2110
Db 987 TGTGTGCTGGGCGGCGCTCTGCTCAGGATGTTGGGACCCAGCTTGTCTCGG 1043
QY 2111 CAGCTAAGAGCAGTACCAGGATGATGATTTGGGACCTGTGTGG 2157
Db 1044 -MCHAAARAGCAGTNACMA-GATGTGATTTGGGACTGTGTGG 1088

RESULT 5
BX360491/c
LOCUS
DEFINITION
BX360491 Homo sapiens PLACENTA cot 25-NORMALIZED Homo sapiens cDNA
clone CS0D1070YF10 3-PRIME, mRNA sequence.
ACCESSION
BX360491
VERSION
BX360491.1 GI:30374437
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1065)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 10409.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1070D05NPI&cluster=10409.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0D1070D05NPI.

```

```

FEATURES
source
Location/Qualifiers
1..1065
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1070YF10"
/tissue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA cot 25-NORMALIZED"
/notes="First strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 251 a 270 c 300 g 218 t 26 others
ORIGIN
Query Match 36.58; Score 932.8; DB 13; Length 1065;
Best Local Similarity 94.18; Pred. No. 48-209;
Matches 1000; Conservative 18; Mismatches 38; Indels 7; Gaps 5;
QY 1468 CAAGTCCAGAGTGCCTGTCTCCGTGAGAAACATCAAGCAGAGAAACAGAGCACTGAGCCT 1527
Db 1057 CAAGTCCAGAGTGCCTGTCTCCGTGAGAAACATCAAGCAGAGAAACAGAGCACTGAGCCT 1000
QY 1528 TCCTGCCCTCCAGCTGCTGCTTCTAGCTGGAAGCCAGCCAGCCTTCTCTACCAACCCAC 1587
Db 999 TCCTGCCCTCCAGCTGCTGCTTCTAGCTGGAAGCCAGCCAGCCTTCTCTACCAACCCAC 942
QY 1588 CATGCCCTCCACCTGACTTTGGCTTGAGAGTATCTCTCTGTGTAATTTCTCCCGGTG 1647
Db 941 CATGCCCTCCACCTGACTTTGGCTTGAGAGTATCTCTCTGTGTAATTTCTCCCGGTG 882
QY 1648 CTGTGAAGGCTGACCGGTGAGAGCCTGCTGGGCTCTCTGGGACCCGCTGTTGCTTCT 1707
Db 881 CTGTGAAGGCTGACCGGTGAGAGCCTGCTGGGCTCTCTGGGACCCGCTGTTGCTTCT 824
QY 1708 GCCTCTCCCTGTGAAAGTCTATATGACGGCGCGCTGAGGCCCGCCAGAACTCGTCTG 1767
Db 823 GCCTCTCCCTGTGAAAGTCTATATGACGGCGCGCTGAGGCCCGCCAGAACTCGTCTG 764
QY 1768 AACCACTTTTCCAGCCAGAGTTCCCAAAGCTGGAACGCTAGCTGCTCTCTCTCTAA 1827
Db 763 AACCACTTTTCCAGCCAGAGTTCCCAAAGCTGGAACGCTAGCTGCTCTCTCTCTAA 704
QY 1828 GATGGCTCTCCCGCAGCCGCGCCCTGAGTTCGAGGATGGGGCCAGCAGCTGTCA 1887
Db 703 GATGGCTCTCCCGCAGCCGCGCCCTGAGTTCGAGGATGGGGCCAGCAGCTGTCA 644
QY 1888 CACTGTGGAATCAAGACAGTGAATCTCTCTGCTGCTGAAACAGTCAATGTAATTAAGTTC 1947
Db 643 CACTGTGGAATCAAGACAGTGAATCTCTCTGCTGCTGAAACAGTCAATGTAATTAAGTTC 584
QY 1948 TAGAGCAGCTCTCTGAGCAGGATTAAGGTCCTGAGAGTGTGTGTGGTGGGGCAGC 2007
Db 583 TAGAGCAGCTCTCTGAGCAGGATTAAGGTCCTGAGAGTGTGTGTGGTGGGGCAGC 524
QY 2008 CTCTGCTCAAAAATTCACCAAGCAGATGCTCTCAGCTCAGCTCATGTGTGGTCTCTGCT 2067
Db 523 CTCTGCTCAAAAATTCACCAAGCAGATGCTCTCAGCTCAGCTCATGTGTGGTCTCTGCT 464
QY 2068 CTCTGCTCAAAAATTCACCAAGCAGATGCTCTCAGCTCAGCTCATGTGTGGTCTCTGCT 2127
Db 463 CTCTGCTCAAAAATTCACCAAGCAGATGCTCTCAGCTCAGCTCATGTGTGGTCTCTGCT 404
QY 2128 CCAGGATGGAATTTTGGGACCTGCTGTGGTGGCTTCTGAGCTGCTTCTGTGTTGTGAG 2187
Db 403 CCAGGATGGAATTTTGGGACCTGCTGTGGTGGCTTCTGAGCTGCTTCTGTGTTGTGAG 344
QY 2188 GACTGACTCCCAATTTCCCTAAAGGAAATGCCCGGGGAGGACATTTGGGAGGAAGATGCC 2247
Db 343 GACTGACTCCCAATTTCCCTAAAGGAAATGCCCGGGGAGGACATTTGGGAGGAAGATGCC 284
QY 2248 TGAGTGTGAGCTTTGGCTCTGCTGCTGCTGAGCCCGCTGAAAGATTAATTCATCA 2307

```

```

Db      283 TGAGTGTGCACTTGGCTGTGCTACCTGCTCTGTAAGAGCCCGCTAAAAATAATCATCCA 224
Qy      2308 AGATTCCTTTTGTAGTTAAAGGGTCCAGTTCTGACTGGAGCCTCTAGAGAGCTGGGCTGTG 2367
Db      223 AGATTCCTTTTGTAGTTAAAGGGTCCAGTTCTGACTGGAGCCTCTAGAGAGCTGGGCTGTG 164
Qy      2368 ATGTTCTTTTGGCTTTTGTCTTACTTAAATGAAGAAACCATGCTGGAGGGCGGTGA 2427
Db      163 ATGTTCTTTTGGCTTTTGTCTTACTTAAATGAAGAAACCATGCTGGAGGGCGGTGA 104
Qy      2428 ACACAGAACCCCTCAAGACAAGATGACAGAGTGGAGGACACATCTAGTGCCTCCATTGCAA 2487
Db      103 ACACAGAACCCCTCAAGACAAGATGACAGAGTGGAGGACACATCTAGTGCCTCCATTGCAA 44
Qy      2488 CC-TGACTGGGCTCCCGGACTCTGTGTGTGAGAAATTAACC 2529
Db      43 CCNNCACTGGGCTCCGGGACTGTGTGTCTNNNDKTRRACS 1

RESULT 6
AL560108
LOCUS
DEFINITION
AL560108 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
clone CSODG002YE14 5-PRIVE, mRNA sequence.
ACCESSION
AL560108
VERSION
AL560108.2 GI:31284239
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 15, 2001 this sequence version replaced gi:12906251.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 10409.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODG002BC07QP1&cluster=10409.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CSODG002BC07QP1.
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODG002YE14"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 259 a 294 c 313 g 288 t 47 others
ORIGIN
Query Match 34.4%; Score 878; DB 9; Length 1201;
Best Local Similarity 99.8%; Pred. No. 3.4e-196;
Matches 900; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy      1643 CGGTGCTGTGAGAGCTGGAGCTGGAGACCTGCTGGGGTCTCTGGGACCCCGCTGTG 1702
Db      68 CGGTGCTGTGAGAGCTGGAGCTGGAGACCTGCTGGGGTCTCTGGGACCCCGCTGTG 127

```

```

Qy      1703 CTTCTGCCCTCCCTGTGGAAGGTCTATATGACGGCGCCCTGAGGCCCCAGAACTCGT 1762
Db      128 CTTCTGCCCTCCCTGTGGAAGGTCTATATGACGGCGCCCTGAGGCCCCAGAACTCGT 187
Qy      1763 CTGTGAACCACTTTTTCAGCCAGAGTTCCAAAGCTGGAACCGCTAGCTGCCTGCTCTTC 1822
Db      188 CTGTGAACCACTTTTTCAGCCAGAGTTCCAAAGCTGGAACCGCTAGCTGCCTGCTCTTC 247
Qy      1823 CTTAAGATGCTCCCTCCCGGACCGCGGACCGCCCTCAGTTGCCAGGATGGGGCCACCAC 1882
Db      248 CTTAAGATGCTCCCTCCCGGACCGCGGACCGCCCTCAGTTGCCAGGATGGGGCCACCAC 307
Qy      1883 TGTCACTCTGTGAATACAGACAGTGAACCTGTCTGCTGAAACGAGTCATCTAATTA 1942
Db      308 TGTCACTCTGTGAATACAGACAGTGAACCTGTCTGCTGAAACGAGTCATCTAATTA 367
Qy      1943 AGTTCTAGAGCAGCTCTCTGAGCAGGATAGGTCCCTGACAGTGAAGTGTGTGTGGG 2002
Db      368 AGTTCTAGAGCAGCTCTCTGAGCAGGATAGGTCCCTGACAGTGAAGTGTGTGTGGG 427
Qy      2003 GCAGCCTCTGCTCAAAAATTACCAAGCAGATGCTCTCAGCCCTCATGTGTTGCTCT 2062
Db      428 GCAGCCTCTGCTCAAAAATTACCAAGCAGATGCTCTCAGCCCTCATGTGTTGCTCT 487
Qy      2063 CTGCTCCTCTAGCTCCCGAGGATGTTGGGACCCAGCTTGTCTGGCAGCTAAGAAC 2122
Db      488 CTGCTCCTCTAGCTCCCGAGGATGTTGGGACCCAGCTTGTCTGGCAGCTAAGAAC 547
Qy      2123 AGTGACACAGATGAGATTTGGCGACCTGTGTGGGCTTGGCTTGTGTTGTTT 2182
Db      548 AGTGACACAGATGAGATTTGGCGACCTGTGTGGGCTTGGCTTGTGTTGTTT 607
Qy      2183 GTGAGGACTGACTCCCATTTCTTAAAGGAAATGCCCGGGGAGGACATTTGGGAGAGA 2242
Db      608 GTGAGGACTGACTCCCATTTCTTAAAGGAAATGCCCGGGGAGGACATTTGGGAGAGA 567
Qy      2243 TGGCCTGAGTGTGCACTTTGGCTTGTCTTACCTGCTCTGAAACCCCGCTTAAATAATTC 2302
Db      568 TGGCCTGAGTGTGCACTTTGGCTTGTCTTACCTGCTCTGAAACCCCGCTTAAATAATTC 727
Qy      2303 ATCCAAGATTCCTTTGTAGTTAAAGGGTCCAGTTCTGACTGGAGCCTCTAGAGAGCTGG 2362
Db      728 ATCCAAGATTCCTTTGTAGTTAAAGGGTCCAGTTCTGACTGGAGCCTCTAGAGAGCTGG 787
Qy      2363 CTTGTATGTTCTTTTGGCTTTTGTCTTAAATGAAGAAACCATGCTGGAGGGGC 2422
Db      788 CTTGTATGTTCTTTTGGCTTTTGTCTTAAATGAAGAAACCATGCTGGAGGGGC 847
Qy      2423 CGTGAACACAGAACCCCTCAAGACAAGATGACAGAGTGGAGGACACATCTAGCTGCCAT 2482
Db      848 CGTGAACACAGAACCCCTCAAGACAAGATGACAGAGTGGAGGACACATCTAGCTGCCAT 907
Qy      2483 TGCACCTCTAGTGGGCTCCCGAGACTCTGTGTGTGAGAAATTAACCCCTGCTGTGTTG 2542
Db      908 TGCACCTCTAGTGGGCT-CCAGACTCTGTGTGTGAGAAATTAAC-CCCTGCTGTGTTG 965
Qy      2543 AG 2544
Db      966 AG 967

RESULT 7
BM909451
LOCUS
DEFINITION
AGENCOURT_6640956 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5434030
5', mRNA sequence.
ACCESSION
BM909451
VERSION
BM909451.1 GI:19359830
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```


laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC Library."

BASE COUNT 194 a 286 c 249 g 171 t

ORIGIN

Query Match 31.1%; Score 794.6; DB 13; Length 900;
Best Local Similarity 98.2%; Pred. No. 1.6e-176;
Matches 814; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 1189 AGTCTACCGCTACAGTCCAGCTCTCTGTGCGCAGAACCGGCACACCGCTGAAGG 1248
DB 1 AGTCTACCGCTACAGTCCAGCTCTCTGTGCGCAGAACCGGCACACCGCTGAAGG 60

QY 1249 CCAAGAGGCGCAAGAGAGAAAGAAAGAGTTGGAGAAAGAAAGAAAGAAATCCCGC 1308
DB 61 CCAAGAGGCGCAAGAGAGAAAGAAAGAGTTGGAGAAAGAAAGAAAGAAATCCCGC 120

QY 1309 CCACTAAGAGCGCACCGCTCAGACAGGGGTCCAGAGAGCCCTGCTGGAGGACGACCTC 1368
DB 121 CCACTAAGAGCGCACCGCTCAGACAGGGGTCCAGAGAGCCCTGCTGGAGGACGACCTC 180

QY 1369 AGGCTGCCAGAGATCTGTGCGAGCTGTTCCTGGCGAGATCAATGCTGTGCTGTGA 1428
DB 181 AGGCTGCCAGAGATCTGTGCGAGCTGTTCCTGGCGAGATCAATGCTGTGCTGTGA 240

QY 1429 AGAAGAGCATCTAGACGTGGCTCCCGCAGCAAGGCTTCAAGTCCAGAGTGCCTGCT 1498
DB 241 AGAAGAGCATCTAGACGTGGCTCCCGCAGCAAGGCTTCAAGTCCAGAGTGCCTGCT 300

QY 1489 CGTCCAGAACTCAAGCAGGAAACAGACGATGAGCTTCTGCTCCAGCTGGCTTC 1548
DB 301 CGTCCAGAACTCAAGCAGGAAACAGACGATGAGCTTCTGCTCCAGCTGGCTTC 360

QY 1549 TAGCTGGAAGCGAGCGCTTCTTACACACACACACACCTGCTCCAGCTGATTTGG 1608
DB 361 TAGCTGGAAGCGAGCGCTTCTTACACACACACACACCTGCTCCAGCTGATTTGG 420

QY 1609 CTTGAGAGTGAATCTCTGTGTAATTTCTGCCCGGTGTGTGAAGCTGGACGGTGA 1668
DB 421 CTTGAGAGTGAATCTCTGTGTAATTTCTGCCCGGTGTGTGAAGCTGGACGGTGA 480

QY 1669 GAGCTGTGGGCTCTCTGGAGCGGCTGTGCTTCTGCTCCCTCTGTGGAAGGTC 1728
DB 481 GAGCTGTGGGCTCTCTGGAGCGGCTGTGCTTCTGCTCCCTCTGTGGAAGGTC 540

QY 1729 TATATGACGGCGCGCTGAGGCGCCAGAACTCGTGTGTAACCACTTTTCCAGCAGAG 1788
DB 541 TATATGACGGCGCGCTGAGGCGCCAGAACTCGTGTGTAACCACTTTTCCAGCAGAG 600

QY 1789 TTCCCAAGAGCTGGAAGCTAGTGTGCTGCTTCTTAAAGATGGCTCCCGGACCGC 1848
DB 601 TTCCCAAGAGCTGGAAGCTAGTGTGCTGCTTCTTAAAGATGGCTCCCGGACCGC 660

QY 1849 CACGGCCCTCAGTTCGCCAGGATGGGCGCACACTGTGCACACTGTGAATACAGACAGT 1908
DB 661 CACGGCCCTCAGTTCGCCAGGATGGGCGCACACTGTGCACACTGTGAATACAGACAGT 720

QY 1909 GAACCTGTCTGCTGGAACGAGTGAATTAAGTTCTAGACAGCTCTCTGAGCAGG 1968
DB 721 GAACCTGTCTGCTGGAACGAGTGAATTAAGTTCTAGACAGCTCTCTGAGCAGG 780

QY 1969 ATAA-GGTCCCTGACAGTGTGTGGTGGGGGAGCTCTGCTC 2016
DB 781 ATAA-GGTCCCTGACAGTGTGTGGTGGGGGAGCTCTGCTC 829

RESULT 9
BQ927459
LOCUS
DEFINITION AGENCOURT_882242 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6381422
5', mRNA sequence.

BQ927459
BQ927459.1 GI:22342490
EST.
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 927)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10M2571 row: j column: 15
High quality sequence stop: 620.

FEATURES
source
1..927
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6381422"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 40"
/notes="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; CDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH MGC Library." 2 others

BASE COUNT 194 a 254 c 251 g 226 t

ORIGIN

Query Match 30.6%; Score 783; DB 13; Length 927;
Best Local Similarity 96.9%; Pred. No. 8.6e-174;
Matches 809; Conservative 0; Mismatches 22; Indels 4; Gaps 1;

QY 1625 CTCTGTGTAATTTCTGCCCGGTGTGTGAAGCTGACGCTGGAGACCTGCTGGGCTCT 1684
DB 1 CTCTGTGTAATTTCTGCCCGGTGTGTGAAGCTGACGCTGGAGACCTGCTGGGCTCT 60

QY 1685 CTTGGACCGCGCTGTGCTTCTGCCCTCCCTGTGGAAGGTCTATATGACGGCGCGCC 1744
DB 61 CTTGGACCGCGCTGTGCTTCTGCCCTCCCTGTGGAAGGTCTATATGACGGCGCGCC 120

QY 1745 TGAGGCCCCAGAACCTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCTCAAGCTGGAAC 1804
DB 121 TGAGGCCCCAGAACCTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCTCAAGCTGGAAC 180

QY 1805 GCTAGCTGCTCTCTTCTTAAAGTTCCTAGACAGCTCTCTGAGCAGGATAGTCCCTGACA 1864
DB 181 GCTAGCTGCTCTCTTCTTAAAGTTCCTAGACAGCTCTCTGAGCAGGATAGTCCCTGACA 240

QY 1865 CAGGATGGGCGCCACCTGTGCATCTGTGAATACAGACAGTGAATCTGTCTGCTG 1924
DB 241 CAGGATGGGCGCCACCTGTGCATCTGTGAATACAGACAGTGAATCTGTCTGCTG 300

QY 1925 AACGAGTCAATGTAATTAAGTTCCTAGACAGCTCTCTGAGCAGGATAGTCCCTGACA 1984
DB 301 AACGAGTCAATGTAATTAAGTTCCTAGACAGCTCTCTGAGCAGGATAGTCCCTGACA 360

QY 1985 GTGAGTGTGGTGGGCGCAGCTCTGCTCAAAATTCACCAAGCAGATGCTCTCA 2044
DB 361 GTGAGTGTGGTGGGCGCAGCTCTGCTCAAAATTCACCAAGCAGATGCTCTCA 420


```

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DMP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1302 row: 1 column: 09
High quality sequence stop: 547.
FEATURES
source
1..876
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6050456"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 72"
/note="Organ: Skin; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 211 a 246 c 253 g 164 t 2 others
ORIGIN
Query Match 29.1%; Score 743.2; DB 13; Length 876;
Best Local Similarity 96.7%; Pred. No. 2.1e-164;
Matches 812; Conservative 0; Mismatches 20; Indels 8; Gaps 5;
QY 495 GAGCACTCTGTGGCCACATGTGGAGCCCTGTGATCCACCCAGTGGTGTGTCAGAGGT 554
Db 1 GAGCACTCTGTGGCCACATGTGGAGCCCTGTGATCCACCCAGTGGTGTGTCAGAGGT 60
QY 555 GTGGGCCACGGCATCAAGATGTGACAGAGCGGCAAGTGCATGCCCATCACCCACCTTC 614
Db 61 GTGGGCCACGGCATCAAGATGTGACAGAGCGGCAAGTGCATGCCCATCACCCACCTTC 120
QY 615 CGGAAGATCTAGTGTGATCGCGTTCCTTACCTCTTCAAGAGGTACGAGCATCTACACA 674
Db 121 CGGAAGATCTAGTGTGATCGCGTTCCTTACCTCTTCAAGAGGTACGAGCATCTACACA 180
QY 675 GAAGCGGTGTGTTGAGGAAGCGATGAAGCTGGAGGCCATTGACCCCTGAATCTGGGC 734
Db 181 GAAGCGGTGTGTTGAGGAAGCGATGAAGCTGGAGGCCATTGACCCCTGAATCTGGGC 240
QY 735 ACACTGTGCGTGGCACTCTGTGAAGTTCCTCTGATGATGATGATGATGATGATGATG 794
Db 241 ACACTGTGCGTGGCACTCTGTGAAGTTCCTCTGATGATGATGATGATGATGATGATG 300
QY 795 GACGGGGGGCCCTCCACAGATGGTGGAGCTGTTCTGCTACCATGCTCTTCCACGCC 854
Db 301 GACGGGGGGCCCTCCACAGATGGTGGAGCTGTTCTGCTACCATGCTCTTCCACGCC 360
QY 855 ATCTTCCCGGCCACCTTCTGTGAGAGAAATGACATTTAGCTCAGACCCGCAAAAGTTAT 914
Db 361 ATCTTCCCGGCCACCTTCTGTGAGAGAAATGACATTTAGCTCAGACCCGCAAAAGTTAT 420
QY 915 GAGGCACAGACTTCAACTGGAGAACTACTTGGAGAGACCAAGTGAAGCGCGTCCA 974
Db 421 GAGGCACAGACTTCAACTGGAGAACTACTTGGAGAGACCAAGTGAAGCGCGTCCA 480
QY 975 TCAGACTCTTTACATGGAATGCCCAACCAATGGCTTCAAGTGGGATGAAGCTGGAG 1034

```

```

Db 481 TCAGACTCTTTACATGGAATGCCCAACCAATGGCTTCAAGTGGGATGAAGCTGGAG 540
QY 1035 GCGTGTGACCTGTGATGAGAGCCCGCGCTCATCTGTGTGCCACCGTGAACAGTGTGCAT 1094
Db 541 GCGTGTGACCTGTGATGAGAGCCCGCGCTCATCTGTGTGCCACCGTGAACAGTGTGCAT 600
QY 1095 GCGTGTCTCAGCATCCACTTTTGACGGCTGGGACAGCGAGTACGACAGTGGTGGACTGC 1154
Db 601 GCGTGTCTCAGCATCCACTTTTGACGGCTGGGACAGCGAGTACGACAGTGGTGGACTGC 660
QY 1155 GAGTCCCGACAGATCTACCCCGCTGGGTGTGTGAGCTCACC-GGTTACAGTCTCC-AGC 1212
Db 661 GAGTCCCGACAGATCTACCCCGCTGGGTGTGTGAGCTCACCAGCGCTTACCAGTCTCCAAGC 720
QY 1213 CTCCTGTGGCGGAG-AACCGGCCACACCTCTGAAGG--CCAAAGAGGCCACAAAGAGA 1269
Db 721 CTCCTGTGGCGGAGAAACCGGCCACACCTCTGAAGGCGCCAAAGAGGCGCCCNNAAGA 780
QY 1270 AAAAGAAACAGTTTGGG---AAGAAAAGAAAAGAAATCCCGCCCACTAAGACGCGACCCC 1326
Db 781 AAAAGAAACCGTTTGGGAAAGAAAAGAAAATCCCGCCCACTAGGAAGCGGACCC 840

RESULT 12
BI092956 851 bp mRNA linear EST 20-JUN-2001
LOCUS 602857991F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4999294 5',
DEFINITION mRNA sequence.
ACCESSION BI092956
VERSION BI092956.1 GI:14511286
KEYWORDS ESR.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1029 row: e column: 23
High quality sequence stop: 803.
FEATURES
source
1..851
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4999294"
/cell_line="MGC36"
/lab_host="DH10B"
/clone_lib="NIH_MGC 10"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5 kb. Library prepared by Life
Technologies."
BASE COUNT 182 a 234 c 231 g 204 t
ORIGIN
Query Match 29.0%; Score 739.8; DB 12; Length 851;
Best Local Similarity 97.5%; Pred. No. 1.3e-163;
Matches 836; Conservative 0; Mismatches 12; Indels 9; Gaps 8;
QY 1668 AGGACCTGCTGGGGTCTCCTGGGAGCCCGCTGCTGCTTCTGCTCCCTCCCTGTTGGAAGGT 1727

```

Db	1	AGGACCTGCTGGGGTCTCCTGGGACCGCGCTGTTGCTTCTGCCCTCCCTGTGGAAGGT	60
Qy	1728	CTATATGACGGGCGCGCTGAGGCCCCAGAACTCGTCTGTGTAACACACCTTTCCAGCCAGA	1787
Db	61	CTATATGACGGGCGCGCTGAGGCCCCAGAACTCGTCTGTGTAACACACCTTTTCAGGCAGA	120
Qy	1788	GTTCCTCAAAGCTGGAAACGCTAGTGCCTGTCTTTCCTTAAGATGGCTCCCCCGACCCG	1847
Db	121	GTTCCTCAAAGCTGGAAACGCTAGTGCCTGTCTTCTTAAATAGTGGCTCCCCCGACCCG	180
Qy	1848	CCACGGGCCCTCAGTTGCTACAGGGATGGGGCCACACACTGTCTACACTGTGGATACAAACAG	1907
Db	181	CCACGGGCCCTCAGTTGCCA--GGATGGGGCCACACACTGTCTACACTGTGGATACAAACAG	239
Qy	1908	TGAACCTCTGTCTGCCGTGAACGAGTCATGTAAATTAAGTTCTAGACAGCTCTCTGAGCAG	1967
Db	240	TGAACCTGTCTCTCCTGAACGAGTCATGTAAATTAAGTTCTAGACAGCTCTCTGAGCAG	299
Qy	1968	GATAAGTCCCTGACAGATGAGTTGTGTGGTGGGGCAGCCTCTGGCTCAAAATTCACC	2027
Db	300	GATAAGTCCCTGACAGATGAGTTGTGTGGTGGGGCAGCCTCTGGCTCAAAATTCACC	359
Qy	2028	AAGCAGAATGCCCTCTCAGCCCTCATGTGTGGTGGCTCTGTCTCCTCTAGCTCCCGAGGAT	2087
Db	360	AAGCAGAATGCCCTCTCAGCCCTCATGTGTGGTGGCTCTGTCTCCTCTAGCTCCCGAGGAT	419
Qy	2088	GTTGGGGAACGACTGTCTCGCAGCTAAGAACAGTGAACAGAGTGTGGATTTTGGCG	2147
Db	420	GTTGGGGAACGACTGTCTCGCAGCTAAGAACAGTGAACAGAGTGTGGATTTTGGCG	479
Qy	2148	ACCTGTGTGGTGGCCCTTGAGCTGCTTCTGTGTGTGTGAGGACTGACTCCCATTTCTTAA	2207
Db	480	ACCTGTGTGGTGGCCCTTGAGCTGCTTCTGTGTGTGTGAGGACTGACTCCCATTTCTTAA	539
Qy	2208	AGGAATGCCCCGGGGAGGACATTTGGGAGGAAGATGGCCTGAGTGTGCATTTGGCTCT	2267
Db	540	AGGAATG--CCCCGGGGAGGACATTTGGGAGGAAGATGGCCTGAGTGTGCATTTGGCTCT	598
Qy	2268	GCTACTGCTCTGTAAGCCCCGCTTAA--NATTAATTCATCCAGATTCCTTTGTAGTTAAA	2326
Db	599	GCTACTGCTCTGTAAGCCCCGCTTAAACAAATTAATTCATCCAGATTCCTTTGTAGTTCAA	658
Qy	2327	GGGT--CCAGTTCTGACTGGAGCCTCTAGAGACTGGGCTTGTATGTTCTTTTGGCTTTT	2385
Db	659	GGGTCCAGTTCTGATTGGAGCCTCTAGAGACTGGGCTTGTATGTTCTTTTGGCTTAG	717
Qy	2386	GTTCCTACTTAATGAAGAAACCATGCTGGAGGGGCGGTGAACACAGAACCCCTCAAGAC	2445
Db	718	GTTCCTACTTAATGAAGAAAGCATGCTGGAGGGGCGGTGAACACAGAACCCCTCAAGAC	777
Qy	2446	AAGGATGA--CAGAGCTGGAGGACACATCTAGCTGCGCATTCGAACCTCACTGGGCTCCCCA	2504
Db	778	AAGGATGACCAGAGCTGCAGGACACAT--TACTGCCAT--GCAGCTCACTGGGCTCCCCA	834
Qy	2505	GACTCTGTGTGAGAA	2521
Db	835	GACTCTGTGTGAGAA	851

RESULT 13

EG396658	EG396658	1057 bp	mRNA	linear	EST 12-MAR-2001
LOCUS	G02459736F1	NIH_MGC_16	Homo sapiens	cdna clone IMAGE:4581846	5',
DEFINITION	mRNA sequence.				
ACCESSION	EG396658				
VERSION	EG396658.1	GI:13290106			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 1057)				
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .				

```

TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-re@mail.nih.gov
           Tissue Procurement: ATCC
           CDNA Library Preparation: Ling Hong/Rubin Laboratory
           CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
           Plate: LLCW1303 row: d column: 07
           High quality sequence stop: 763.
           Location/Qualifiers
             1..1057
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="IMAGE:4581846"
               /tissue_type="retinoblastoma"
               /lab_host="DH10B (phage-resistant)"
               /clone_lib="NIH_MGC_16"
               /note="Organ: eye; Vector: pOTB7; Site: 1: XhoI; Site 2:
               EcoRI; CDNA made by oligo-dT priming. Directionally
               cloned into EcoRI/XhoI sites using the following 5',
               adaptor: GGCACGAG(G). Library constructed by Ling Hong
               in the laboratory of Gerald M. Rubin (University of
               California, Berkeley) using ZAP-CDNA synthesis kit
               (Stratagene) and Superscript II RT (Life Technologies).
               Note: this is a NIH MGC Library."
           254 a 296 c 303 g 203 t
BASE COUNT

```

BASE COUNT	254 a	296 c	303 g	203 t	1 others
ORIGIN	Note: this is a NIH/MGC Library. "				
Query Match	28.9%	Score 737.6;	DB 10;	Length 1057;	
Best Local Similarity	93.4%	Prod No. 4.5e-163;			
Matches 825,	Conservative 0;	Mismatches 50;	Indels 8;	Gaps 5;	
Qy	1247	GGCCAAAGAGGCCCAAAAGAGAAAGAAAAGAACACAGTTTGGGAAGAAAAGGAAAAGAAATCCC	1306		
Db	1	GGCCAAAGAGGCCCAAAAGAGAAAGAAAAGAGACAG-ATTGGGAAGAAAAGGAAAAGAAATCCC	59		
Qy	1307	GCCCACTAAGACGCGACCCCTCAGACAGGGGTCCCAAGAACCCCTGCTGGAGACGACCC	1366		
Db	60	GCCCACTAAGACGCGACCCCTCACACAGGGGTCCAGAACCCCTGCTGGAGACGACCC	119		
Qy	1367	TCAGGGTGCAGGAAGATCTCGTCGGAGCCTGTTCTCTGGCGAGATCATTCGTGTGCCTGT	1426		
Db	120	TCAGGGTGCAGGAAGATCTCGTCGGAGCCTGTTCTCTGGCGAGATCATTCGTGTGCCTGT	179		
Qy	1427	GAAGGAAGACATCTAGACGTGSCCTCGCCGACAAAGGCTTCAAGTCCAGAGCTGCCTGT	1486		
Db	180	GAAGGAAGACATCTAGACGTGSCCTCGCCGACAAAGGCTTCAAGTCCAGAGCTGCCTGT	239		
Qy	1487	CTCCGTCGAGAACATCAAGCAGGAAACAGACGACTGAGCCTTCCTGCCTCCAGCCTGGCT	1546		
Db	240	CTCCGTCGAGAACATCAAGCAGGAAACAGACGACTGAGCCTTCCTGCCTCCAGCCTGGCT	299		
Qy	1547	TCCTAGCTGGAAGCCAGCCAGACGTTTTCTTACCACCAACACCATGCTCCACCTGACTTT	1606		
Db	300	TCCTAGCTGGAAGCCAGCCAGACGTTTTCTTACCACCAACACCATGCTCCACCTGACTTT	359		
Qy	1607	GGCTTGGAGACTGATCCTCTCTGTGTAAATTCCTGCCCCGCTGTGTGAAGCTCGACGGTG	1666		
Db	360	GGCTTGGAGACTGATCCTCTCTGTGTAAATTCCTGCCCCGCTGTGTGAAGCTCGACGGTG	419		
Qy	1667	GAGGACCTGCTGGGGTCTCCTGGACCCGCCCTGTGTTCTGTCCTCCCTCGCTGGAAAGG	1726		
Db	420	GAGGACCTGCTGGGGTCTCCTGGACCCGCCCTGTGTTCTGTCCTCCCTCGCTGGAAAGG	479		
Qy	1727	TCTATATGACGGCGCGCTGAGGCCCCAGAACTCGTCTGTGAACACACCTTTTCCAGCCAG	1786		
Db	480	TCTATATGACGGCGCGCTGAGGCCCCAGAACTCGTCTGTGAACACACCTTTTCCAGCCAG	539		
Qy	1787	AGTTTCCAAAGCTGGAACGCTAGCTGCTGTCTTCTCTTAAAGATGSCCTCCCCCCCCACCC	1846		

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 4, 2004, 13:26:23 ; Search time 673 Seconds
(without alignments)
10248.238 Million cell updates/sec

Title: US-10-031-915-90
Perfect score: 2555
Sequence: 1 agcagtcctccactatgac.....ttgcttgagaaaaa 2555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2555	100.0	2555	AAFS9625	Human cell cycle a
2	2550.8	99.8	2781	AA158732	Human polynucleoti
3	2544	99.6	3530	ACG50583	Human secreted pro
4	2544	99.6	3530	ABZ71320	Secreted protein-e
5	2514.4	98.4	2663	24 ABL99857	Human polynucleoti
6	2374.2	92.9	2380	22 AAI60518	Human polynucleoti
7	1434.6	56.1	3071	24 ABA93718	Human transmembran
8	413.8	16.2	3590	21 AA294124	Haematopoietic ste

9	405.2	15.9	2564	25	ABX08841	Angiogenesis-assoc
10	324.4	12.7	476	24	ABL77723	Human ovarian canc
11	281.4	11.0	418	24	ABL68929	Kidney cancer rela
12	271.8	10.6	289	20	AAH87128	Human single nucle
13	266.4	10.4	3486	23	ABL27733	Drosophila melanog
14	266.4	10.4	4379	23	ABL19885	Drosophila melanog
15	252.8	9.9	290	16	AAT23623	Human gene signatu
16	250.6	9.8	251	19	AAI12386	Human biallelic po
17	235.4	9.2	558	22	ABA63686	Human foetal liver
18	235.4	9.2	558	22	ABA30876	Probe #9342 for ge
19	235.4	9.2	558	22	AAK12202	Human brain expres
20	235.4	9.2	558	22	AAK37924	Human bone marrow
21	235.4	9.2	558	22	AAI18685	Probe #8618 for ge
22	235.4	9.2	558	22	AAI43798	Probe #12484 used
23	235.4	9.2	558	23	ABS37546	Human liver single
24	235.4	9.2	558	24	ABS11920	Human genome-deriv
25	232	9.1	232	22	ABA75996	Human foetal liver
26	232	9.1	232	22	ABA40556	Probe #19022 for g
27	232	9.1	232	22	AAK24673	Human brain expres
28	232	9.1	232	22	AAK50672	Human bone marrow
29	232	9.1	232	22	AAI27682	Probe #17615 for g
30	232	9.1	232	22	AAI56650	Probe #25336 used
31	232	9.1	232	23	ABS50269	Human liver single
32	232	9.1	232	24	ABS24144	Human genome-deriv
33	167	6.5	1548	23	ABS78585	DNA encoding novel
34	158.8	6.2	8615	23	ABL27732	Drosophila melanog
35	158.8	6.2	8670	23	ABL19884	Drosophila melanog
36	137.8	5.4	2922	20	AAI27923	Rat l(3)mbl protel
37	135.8	5.3	165	22	AAI25416	Human ovarian PCR-
38	133.6	5.2	3394	20	AAI27920	Human l(3)mbl prot
39	133.6	5.2	3512	20	AAI27921	Human l(3)mbl prot
40	133.4	5.2	4505	25	ABX34736	Human mddt cDNA SE
41	133.4	5.2	4619	25	ABX34823	Human mddt cDNA SE
42	129.4	5.1	2714	20	AAI27922	Mouse l(3)mbl prot
43	121	4.7	555	22	ABA63649	Human foetal liver
44	121	4.7	555	22	ABA30842	Probe #9308 for ge
45	121	4.7	555	22	AAK12168	Human brain expres

ALIGNMENTS

RESULT 1
AAFS9625

ID AAFS9625 standard; cDNA; 2555 BP.

XX AAFS9625;

XX AAFS9625;

XX 24-APR-2001 (first entry)

XX Human cell cycle and proliferation protein CCYPR-36 cDNA, SEQ ID NO:90.

XX Cell cycle and proliferation protein; CCYPR; human; agonist;

XX antgonist; gene therapy; detection; gene therapy;

XX transgenic animal disease model; immune disorder;

XX developmental disorder; cell signalling disorder;

XX cell proliferative disorder; cancer; tumour; anaemia; epilepsy;

XX arteriosclerosis; asthma; allergy; diabetes mellitus;

XX menstrual cycle disorder; bacterial infection; ss.

OS Homo sapiens.

XX WO200107471-A2.

XX 01-FEB-2001.

XX 21-JUL-2000; 2000WO-US19948.

XX 21-JUL-1999; 99US-0145075.

XX 08-SEP-1999; 99US-0153129.

XX 10-NOV-1999; 99US-0164647.

XX (INCY-) INCYTE GENOMICS INC.

XX Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;
PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;
XX WPI: 2001-112727/12.
DR P-PSDB; AAB60488.
XX Human cell cycle and proliferation proteins and polynucleotides are
PT used to treat, diagnose and prevent immune, developmental and cell
PT signaling disorders and cell proliferative disorders including cancer -
XX Claim 5; Page 190-191; 205pp; English.
XX Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human
CC cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506.
CC CCYPR and agonists of CCYPR are used to treat diseases or conditions
CC associated with decreased expression of functional CCYPR, while CCYPR
CC antagonists are used to treat diseases or conditions associated with
CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect
CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)
CC that specifically bind to CCYPR, and in drug screening methods to
CC identify compounds that modulate the activity of CCYPR. CCYPR
CC nucleotides can be used to generate transgenic animal models of human
CC disease, and can be used in gene therapy in target cells with genetic
CC abnormalities with respect to the expression of CCYPR for the
CC treatment or prevention of a disorder associated with CCYPR.
CC Diseases which can be diagnosed, treated and prevented using CCYPR
CC proteins, nucleic acids, agonists or antagonists include immune,
CC developmental and cell signalling disorders, and cell proliferative
CC disorders including cancer. Specific examples of these disorders
CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
CC diabetes mellitus, disorders of the menstrual cycle and infections
CC caused by bacteria.
XX SQ Sequence 2555 BP; 577 A; 703 C; 722 G; 547 T; 0 other;

Query Match 100.0%; Score 2555; DB 22; Length 2555;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGTCCCACTATGACAGTGGAGGAGTGTGATGAAAGGAGGAGTGGAGTGC 60
DB 1 AGCAGTCCCACTATGACAGTGGAGGAGTGTGATGAAAGGAGGAGTGGAGTGC 60

QY 61 TCAACAGTGTGCTGTGCTCCAGCGGGGTGACTGATCGCTCTGTATCCAGACAG 120
DB 61 TCAACAGTGTGCTGTGCTCCAGCGGGGTGACTGATCGCTCTGTATCCAGACAG 120

QY 121 CAGGATATCGGGTGTGCTTCCGATGAGAGGCTTTGAAATGACGCCAGGCATGACTTCT 180
DB 121 CAGGATATCGGGTGTGCTTCCGATGAGAGGCTTTGAAATGACGCCAGGCATGACTTCT 180

QY 181 GGTGCAACCTGGGAACAGTGTGATGTCACCCCATTTGGCTGGTGTGTCATCAACAGCAAGA 240
DB 181 GGTGCAACCTGGGAACAGTGTGATGTCACCCCATTTGGCTGGTGTGTCATCAACAGCAAGA 240

QY 241 TCCTAGTGCCTCCAGGACCATCATGCGCAAGTTCACCGCTGGAAGGGTACCTCATGA 300
DB 241 TCCTAGTGCCTCCAGGACCATCATGCGCAAGTTCACCGCTGGAAGGGTACCTCATGA 300

QY 301 AACGGCTGGTGGTCCAGGACGCTTCCCGTGGATTTCCATCAAGATGGTGGAGAGCA 360
DB 301 AACGGCTGGTGGTCCAGGACGCTTCCCGTGGATTTCCATCAAGATGGTGGAGAGCA 360

QY 361 TGAGTACCCCTTTAGGCAGGGCATGCGGCTGGAGTGGTGGACAGTCCAGGTGTCCAC 420
DB 361 TGAGTACCCCTTTAGGCAGGGCATGCGGCTGGAGTGGTGGACAGTCCAGGTGTCCAC 420

QY 421 GCACCTCGCATGGCTGTGGTGGACACAGTAATCGGGGGTTCGCTACGGCTCTCTACGAGG 480
DB 421 GCACCTCGCATGGCTGTGGTGGACACAGTAATCGGGGGTTCGCTACGGCTCTCTACGAGG 480

QY 481 ATGGTACAGTGCAGCAGTCTTCTGTCGCCATGTGGAGCCCCCTGATCCACCCAGTGG 540
DB 481 ATGGTACAGTGCAGCAGTCTTCTGTCGCCATGTGGAGCCCCCTGATCCACCCAGTGG 540

QY 541 GTTGTACAGTGTGGGCCACGGCATCAGATGTGAGAGGGCAAGTGCATGGGCC 600
DB 541 GTTGTACAGTGTGGGCCACGGCATCAGATGTGAGAGGGCAAGTGCATGGGCC 600

QY 601 ATCACCACACCTTCCGGAAGTCTACTGTGATGCCGTTCCCTTACCTCTTCAAGAGGTAC 660
DB 601 ATCACCACACCTTCCGGAAGTCTACTGTGATGCCGTTCCCTTACCTCTTCAAGAGGTAC 660

QY 661 GAGCAGTCTACACAGAGCGGTTGGTTGAGGAAGGATGAGCTGAGGCGCATTGACC 720
DB 661 GAGCAGTCTACACAGAGCGGTTGGTTGAGGAAGGATGAGCTGAGGCGCATTGACC 720

QY 721 CCTCAATCTGGGCAACATCTGCGTGGCAACTGTCTGTAAGTCTTCTGATGATGATACC 780
DB 721 CCTCAATCTGGGCAACATCTGCGTGGCAACTGTCTGTAAGTCTTCTGATGATGATACC 780

QY 781 TGATGATCTGTGAGACGGGGGCGCTCCACAGATGGCTTGGACTGCTTCTGCTACCATG 840
DB 781 TGATGATCTGTGAGACGGGGGCGCTCCACAGATGGCTTGGACTGCTTCTGCTACCATG 840

QY 841 CCTCTCCACGCGCATCTTCCCGGCACCTTCTGTGCAAGAATGACATTGAGCTCACAC 900
DB 841 CCTCTCCACGCGCATCTTCCCGGCACCTTCTGTGCAAGAATGACATTGAGCTCACAC 900

QY 901 CGCCAAAAGTTATGAGGCACAGACTTTCACTGGGAGAACTACTTTGAGAGAACCAAGT 960
DB 901 CGCCAAAAGTTATGAGGCACAGACTTTCACTGGGAGAACTACTTTGAGAGAACCAAGT 960

QY 961 CGAAAGCGCTCCATCGAGACTCTTTAATCATGGATTGCGCCAAACCATGGCTCAAGGTGG 1020
DB 961 CGAAAGCGCTCCATCGAGACTCTTTAATCATGGATTGCGCCAAACCATGGCTCAAGGTGG 1020

QY 1021 GCATGAAGCTGGAGCGCTGGACCTGATGGAGCCCCGGCTCATCTGTGTGGCCACGGTGA 1080
DB 1021 GCATGAAGCTGGAGCGCTGGACCTGATGGAGCCCCGGCTCATCTGTGTGGCCACGGTGA 1080

QY 1081 AACAGTGGTGTGATCGGCTCTCTAGCATCCATTTGACGGTGGGACAGCGAGTACGACC 1140
DB 1081 AACAGTGGTGTGATCGGCTCTCTAGCATCCATTTGACGGTGGGACAGCGAGTACGACC 1140

QY 1141 AGTGGGTGGACTGCGAGTCCCGAGACATCTACCCGCTGGTGTGTGAGCTCACCGGCT 1200
DB 1141 AGTGGGTGGACTGCGAGTCCCGAGACATCTACCCGCTGGTGTGTGAGCTCACCGGCT 1200

QY 1201 ACCAGCTCCAGCTCTCTGTGGCCGAGAACCGGCCACACCGCTGAAGGCCAAAGAGGCCA 1260
DB 1201 ACCAGCTCCAGCTCTCTGTGGCCGAGAACCGGCCACACCGCTGAAGGCCAAAGAGGCCA 1260

QY 1261 CAAAGAGAAAGAAACAGTTTGGAGAAAGAAAGAAAGAAATCCCGCCCACTAAGAGCC 1320
DB 1261 CAAAGAGAAAGAAACAGTTTGGAGAAAGAAAGAAAGAAATCCCGCCCACTAAGAGCC 1320

QY 1321 GACCCCTCAGACAGAGGGTCCAAAGAGCCCTGTGGAGAGCAGCCCTCAGGGTGCAGGA 1380
DB 1321 GACCCCTCAGACAGAGGGTCCAAAGAGCCCTGTGGAGAGCAGCCCTCAGGGTGCAGGA 1380

QY 1381 AGATCTGTGAGAGCTGTCTTCTGCGGAGATCATTTGCTGTGCGTGTGAAGAGAGCATC 1440
DB 1381 AGATCTGTGAGAGCTGTCTTCTGCGGAGATCATTTGCTGTGCGTGTGAAGAGAGCATC 1440

QY 1441 TAGAGTGTGCTCGCCGACAGGCTTCAAGTCCAGAGCTCGCTGTCTCCGCTCAGAGACA 1500
DB 1441 TAGAGTGTGCTCGCCGACAGGCTTCAAGTCCAGAGCTCGCTGTCTCCGCTCAGAGACA 1500

QY 1501 TCAAGCAGAAACAGACAGCTGAGCTTCTGCTCCAGCTCGGCTTCTAGCTGGAAGCC 1560
DB 1501 TCAAGCAGAAACAGACAGCTGAGCTTCTGCTCCAGCTCGGCTTCTAGCTGGAAGCC 1560

AC	AAI58732;
AD	
DT	22-OCT-2001 (first entry)
XX	
DE	Human polynucleotide SEQ ID NO 935.
XX	
KW	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW	leukaemia; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200153312-A1.
XX	
PD	26-JUL-2001.
XX	
PF	26-DEC-2000; 2000WO-US34263.
XX	
PR	21-JAN-2000; 2000US-0488725.
XX	
PR	25-APR-2000; 2000US-052317.
XX	
PR	09-JUL-2000; 2000US-0598042.
XX	
PR	19-JUL-2000; 2000US-0620312.
XX	
PR	03-AUG-2000; 2000US-0653450.
XX	
PR	14-SEP-2000; 2000US-0682191.
XX	
PR	19-OCT-2000; 2000US-0693036.
XX	
PR	29-NOV-2000; 2000US-0727344.
XX	
PA	(HYSE-) HYSEQ INC..
XX	
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX	
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX	
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX	
DR	WFI; 2001-442253/47.
XX	
DR	P-PSDB; AAM39576.
XX	
PT	Novel nucleic acids and polypeptides, useful for treating disorders
XX	such as central nervous system injuries -
XX	
PS	Claim 1; SEQ ID NO 935; 10078pp; English.
XX	
CC	The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC	the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilisation of the activities such as: Immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	
XX	Sequence 2781 BP; 632 A; 776 C; 784 G; 589 T; 0 other;
XX	
QY	Query Match 99.8%; Score 2550.8; DB 22; Length 2781;
DB	Best Local Similarity 99.9%; Pred. No. 0;
XX	Matches 2552; Conservative 0; Mismatches 2; Indels 0; Gaps 0
QY	1 AGCAGTCCCCTATTATCACCGTCGGAGGATGTGTAAGGGATGAAGGTGGAGTGCG 60
DB	228 AGCAGTCCCCTATTATCACCGTCGGAGGATGTGTAAGGGATGAAGGTGGAGTGCG 287
QY	61 TCACAGTGATGTGTGCTCCGCCAGCGGGTGTACTGGATCGCCTCTGTCATCCAGACAG 120

288 TCAACAGTGTGCTGCTGCCAGCCGGGTGACTGGATCGGCTCTGTCTATCCAGACAG 347
121 CAGGGTATCGGGTCTCTCTTGGTATGAAGGCTTTGAAATGACGCCAGCCATGACTTCT 180
348 CAGGGTATCGGGTCTCTCTTGGTATGAAGGCTTTGAAATGACGCCAGCCATGACTTCT 407
181 GGTCAACCTGGGAACAGTGGATGTCACCCCATTTGGCTGGTGGTGCATCAACAGCAGA 240
408 GGTCAACCTGGGAACAGTGGATGTCACCCCATTTGGCTGGTGGTGCATCAACAGCAGA 467
241 TCCTAGTCCGCCACGACGACCATCCATGCCAAGTTTCCCGACTGGAAGGCTACTCTCATGA 300
468 TCCTAGTCCGCCACGACGACCATCCATGCCAAGTTTCCCGACTGGAAGGCTACTCTCATGA 527
301 AACGGCTGGTGGGCTCAGACAGCTTCCCGTGGATTTCCACATCAAGATGGTGGAGAGCA 360
528 AACGGCTGGTGGGCTCAGACAGCTTCCCGTGGATTTCCACATCAAGATGGTGGAGAGCA 587
361 TGAAGTACCCCTTTAGCAGGGCATCGGCTGCAAGTGGTGGACAGTCCCGAGGTGTAC 420
588 TGAAGTACCCCTTTAGCAGGGCATCGGCTGCAAGTGGTGGACAGTCCCGAGGTGTAC 647
421 GCACTGCATGGCTGTGGTGGACAGTAATCGGGGTCGCTTACCGCTCTCTACGAGG 480
648 GCACTGCATGGCTGTGGTGGACAGTAATCGGGGTCGCTTACCGCTCTCTACGAGG 707
481 ATGGTGACAGTACGACGACTTCTGTGGCCACATGTGGAGCCGCTGTATCCACCCAGTGG 540
708 ATGGTGACAGTACGACGACTTCTGTGGCCACATGTGGAGCCGCTGTATCCACCCAGTGG 767
541 GTTGGTCAAGCATGTGGGCAACGGCATCAAGATGTGAGAGGGGGAAGTGCATGGGCC 600
768 GTTGGTCAAGCATGTGGGCAACGGCATCAAGATGTGAGAGGGGGAAGTGCATGGGCC 827
601 ATCACCCACCTTCCGAGAGTACTGTGATGCCCTTCCCTTACCTTCTTCAAGAGGTTAC 660
828 ATCACCCACCTTCCGAGAGTACTGTGATGCCCTTCCCTTACCTTCTTCAAGAGGTTAC 887
661 GAGCAGTCTACACAGAAGCGGTGGTTTGGAGAGGGATGAAGTGGAGGCCATTGACC 720
888 GAGCAGTCTACACAGAAGCGGTGGTTTGGAGAGGGATGAAGTGGAGGCCATTGACC 947
721 CCTGTAATCTGGCAACATCTGGTGGCACTGTCTGTAAGTTTCTCTGGATGGATACC 780
948 CCTGTAATCTGGCAACATCTGGTGGCACTGTCTGTAAGTTTCTCTGGATGGATACC 1007
781 TGAATGATCTGTGGACGGGGGCCCTCCACAGATGGCTTGGACCTGCTCTACCATG 840
1008 TGAATGATCTGTGGACGGGGGCCCTCCACAGATGGCTTGGACCTGCTCTACCATG 1067
841 CTTCTTCCACGCAATCTTCCGGGCCACTTCTGTGAGAAGATGACATTTGAGCTTCACAC 900
1068 CTTCTTCCACGCAATCTTCCGGGCCACTTCTGTGAGAAGATGACATTTGAGCTTCACAC 1127
901 CGCCAAAAGTTATGAGGCACAGCTTTCACCTGGGAGACTTCTTGGAGAGACCAAGT 960
1128 CGCCAAAAGTTATGAGGCACAGCTTTCACCTGGGAGACTTCTTGGAGAGACCAAGT 1187
961 CGAAAGCCGCTCCATCGAGACTCTTTAATCATGGATTGGCCAAACCATGGCTTCAAGGTGG 1020
1188 CGAAAGCCGCTCCATCGAGACTCTTTAATCATGGATTGGCCAAACCATGGCTTCAAGGTGG 1247
1021 GCATGAAGTGGAGCCGTTGACCTGTATGGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1080
1248 GCATGAAGTGGAGCCGTTGACCTGTATGGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1307
1081 AACGAGTGGTGCATCGGCTCTCAGCATCCACTTTGACGGCTGGACAGCCAGTACGACC 1140
1308 AACGAGTGGTGCATCGGCTCTCAGCATCCACTTTGACGGCTGGACAGCCAGTACGACC 1367
1141 AGTGGGTGATCTGAGTCCCGACATCTACCCCGTCCGGTGGTGTGAGCTCACCGGCT 1200
1368 AGTGGGTGATCTGAGTCCCGACATCTACCCCGTCCGGTGGTGTGAGCTCACCGGCT 1427

1201 ACCAGTCTCCAGCTCTCTGTGSCCGCAGAACCGGCCACACCGCTGAAGGCCAAGAGGCCA 1260
1428 ACCAGTCTCCAGCTCTCTGTGSCCGCAGAACCGGCCACACCGCTGAAGGCCAAGAGGCCA 1487
1261 CAAGAAGAAAAGAAAACAGTTTGGGAAGAAAAGAAAAGAAATCCCGCCCACTAAGACGC 1320
1488 CAAGAAGAAAAGAAAACAGTTTGGGAAGAAAAGAAAAGAAATCCCGCCCACTAAGACGC 1547
1321 GACCCCTCAGACAGGGTCCAGAGAGCCCTGTGTGGAGGACGACCCCTCAGGGTGCAGGA 1380
1548 GACCCCTCAGACAGGGTCCAGAGAGCCCTGTGTGGAGGACGACCCCTCAGGGTGCAGGA 1607
1381 AGATCTCGTCCGAGCCTGTTCTCTGCGCAGATCATTTGCTGTGCGGTGTGAAGAGAGCATC 1440
1608 AGATCTCGTCCGAGCCTGTTCTCTGCGCAGATCATTTGCTGTGCGGTGTGAAGAGAGCATC 1667
1441 TAGACGTGGCTCGCCCGCAGCAAGGCTTCAAGTCCAGAGCTGCCTGTCTCGTCCAGAA 1500
1668 TAGACGTGGCTCGCCCGCAGCAAGGCTTCAAGTCCAGAGCTGCCTGTCTCGTCCAGAA 1727
1501 TCAAGCAGGAAAACAGACGACTGAGCCCTTCTGCTCCAGCCTGGCTTCTAGCTGGAGCC 1560
1728 TCAAGCAGGAAAACAGACGACTGAGCCCTTCTGCTCCAGCCTGGCTTCTAGCTGGAGCC 1787
1561 AGCCAGCGTTTCTTACCACCAACCATGCTCCCTTCCACTGACTTTTGGCTTGGAGACTGA 1620
1788 AGCCAGCGTTTCTTACCACCAACCATGCTCCCTTCCACTGACTTTTGGCTTGGAGACTGA 1847
1621 TCCTCTCTGTGTAATTTCTGCCCGGTGCTGTGAAGCTGGACGGTGGAGGACCTGTGGG 1680
1848 TCCTCTCTGTGTAATTTCTGCCCGGTGCTGTGAAGCTGGACGGTGGAGGACCTGTGGG 1907
1681 GTCTCTGGACCCCGCTGTTGCTTCTGCCCTCCCTGTGGAAGGTCTATATGACGGGC 1740
1908 GTCTCTGGACCCCGCTGTTGCTTCTGCCCTCCCTGTGGAAGGTCTATATGACGGGC 1967
1741 CGCTGAGGCCCCAGAACTGCTGTGAACCACTTTTCCAGCAGAGTTTCCCAAAGCTG 1800
1968 CGCTGAGGCCCCAGAACTGCTGTGAACCACTTTTCCAGCAGAGTTTCCCAAAGCTG 2027
1801 GAACGCTAGTGCCTGCTTCTTCTTAAAGATGGCTCCCGCCAGCCGACAGCCCTCAG 1860
2028 GAACGCTAGTGCCTGCTTCTTCTTAAAGATGGCTCCCGCCAGCCGACAGCCCTCAG 2087
1861 TTGCCAGGGATGGGGCCACCACTGTCACTGTGGAATACAAGACAGTGAACCTCTGCTG 1920
2088 TTGCCAGGGATGGGGCCACCACTGTCACTGTGGAATACAAGACAGTGAACCTCTGCTG 2147
1921 CCTGAACGAGTCAATGTAATTAAGTTCTAGACAGCTCTCTGAGCAGGATTAAGTCCCT 1980
2148 CCTGAACGAGTCAATGTAATTAAGTTCTAGACAGCTCTCTGAGCAGGATTAAGTCCCT 2207
1981 GACAGTGAAGTGTGGTGGGGCAGCTCTGCTTCAAAAATTCACCAAGCAGATGSCCT 2040
2208 GACAGTGAAGTGTGGTGGGGCAGCTCTGCTTCAAAAATTCACCAAGCAGATGSCCT 2267
2041 CTCAGCCTCATGTGTTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2100
2268 CTCAGCCTCATGTGTTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2327
2101 CTTGTCTCGCAGCTAAGAAGCAGTGAACAGGATGGAGTTTGGCGACTGTGTGTGG 2160
2328 CTTGTCTCGCAGCTAAGAAGCAGTGAACAGGATGGAGTTTGGCGACTGTGTGTGG 2387
2161 CTTGTAGCTGCTTCTGTGTTTGTGAGGACTGACTCCCATTTTCTTAAAGAAATGCCGCC 2220
2388 CTTGTAGCTGCTTCTGTGTTTGTGAGGACTGACTCCCATTTTCTTAAAGAAATGCCGCC 2447
2221 GGGAGGACATTTGGGAGGAGATGGCTGAGTGTGCACTTTTGGCTCTGTCTCTCTCTCT 2280
2448 GGGAGGACATTTGGGAGGAGATGGCTGAGTGTGCACTTTTGGCTCTGTCTCTCTCTCT 2507

QY 2281 GAAGCCCGCTAAATATTCATCCAGATTCTTTAGTAAAGGTCAGTTCGA 2340
 Db 2508 GAAGCCCGCTAAATATTCATCCAGATTCTTTAGTAAAGGTCAGTTCGA 2567
 QY 2341 CTGAGCCTCTAGAGAGCTGGCTGTATGTTCTTTGGCTTTTGTCTACCTAAATG 2400
 Db 2568 TTGAGCCTCTAGAGAGCTGGCTGTATGTTCTTTGGCTTTTGTCTACCTAAATG 2627
 QY 2401 AAGAAACCATCCCTGGAGGCGCTGACACAGAACCTCAAGACAGGATGACAGACT 2460
 Db 2628 AAGAAACCATCCCTGGAGGCGCTGACACAGAACCTCAAGACAGGATGACAGACT 2687
 QY 2461 GGAGGACACATCTAGCTGCCATTGCAACCTCATCTGGGCTCCCGAGACTCTGTGTGTAGA 2520
 Db 2688 GGAGGACACATCTAGCTGCCATTGCAACCTCATCTGGGCTCCCGAGACTCTGTGTGTAGA 2747
 QY 2521 AATTAACCCCTCTGCTGTGAGAAAAA 2554
 Db 2748 AATTAACCCCTCTGCTGTGAGAAAAA 2781

RESULT 3
 ACC50583
 ID ACC50583 standard; cDNA; 3530 BP.
 XX
 AC ACC50583;
 XX
 DT 12-JUN-2003 (first entry)
 XX
 DE Human secreted protein coding sequence, SEQ ID 250.
 XX
 KW Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;
 KW vulnerable; antiinflammatory; nootropic; neuroprotective;
 KW antiparkinsonian; gene therapy; human; cardiovascular disorder;
 KW gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WQ200295010-A2.
 XX
 PD 28-NOV-2002.
 XX
 PF 19-MAR-2002; 2002WO-US09785.
 XX
 PR 21-MAR-2001; 2001US-277340P.
 PR 19-JUL-2001; 2001US-306171P.
 PR 13-NOV-2001; 2001US-331287P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SY;
 XX
 DR WPI; 2003-129429/12.
 XX
 PT Novel human secreted proteins, useful for detecting, preventing,
 PT diagnosing, prognosticating, treating and/or ameliorating
 PT cardiovascular disorders such as arrhythmia -
 XX
 PS Claim 21; SEQ ID 250; 1881pp; English.

CC The present invention relates to novel human secreted proteins
 CC (AB47633-AB478145) and their coding sequences (ACC50344-ACC50856). The
 CC proteins and their coding sequences are useful for the preparation of a
 CC diagnostic or pharmaceutical composition for diagnosing or treating a
 CC cardiovascular disorder (e.g., arrhythmia, tachycardia, cardiac arrest,
 CC coronary arteriosclerosis and myocardial ischemia), neural disorders,
 CC immune system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, pulmonary disorders, renal disorders,
 CC proliferative disorders and/or cancerous diseases and conditions, for
 CC wound healing and epithelial cell proliferation, to treat inflammation or
 CC infection, for treating thrombosis and arteriosclerosis, for treating or
 CC preventing neural damage which occurs in neuronal disorders or
 CC neurodegenerative conditions such as Alzheimer's disease and Parkinson's

CC disease, to enhance bone and periodontal regeneration and aid in tissue
 CC transplants or bone grafts, to prevent skin aging or hair loss, to
 CC stimulate growth and differentiation of haematopoietic cells and bone
 CC marrow cells when used in combination with other cytokines, to maintain
 CC organs before transplantation or for supporting cell culture of primary
 CC tissues, to increase or decrease differentiation or proliferation of
 CC embryonic stem cells, or to modulate mammalian characteristics or
 CC metabolism.
 CC Note: The sequence data for this patent was published in electronic
 CC format and is available from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 3530 BP; 815 A; 945 C; 972 G; 790 T; 8 other;
 Query Match 99.6%; Score 2544; DB 25; Length 3530;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGCAGTCCCACTCTATGACCAGTGGGAGGATGATGAAAGGGATGAAGTGGAGTGC 60
 Db 895 AGCAGTCCCACTCTATGACCAGTGGGAGGATGATGAAAGGGATGAAGTGGAGTGC 954
 QY 61 TCACAGTGTATGCTGTCTCCCGAGCGGGTGTACTTGGATCGCTCTGTCTATCCAGACAG 120
 Db 955 TCACAGTGTATGCTGTCTCCCGAGCGGGTGTACTTGGATCGCTCTGTCTATCCAGACAG 1014
 QY 121 CAGGATATCGGGTGTCTGCTTCGGTATGAAGGCTTTGAAATGACGCCAGCCATGACTTCT 180
 Db 1015 CAGGATATCGGGTGTCTGCTTCGGTATGAAGGCTTTGAAATGACGCCAGCCATGACTTCT 1074
 QY 181 GGTCAACCTCGGAAACAGTGTGATGTCACCCATGCTGGTGGTGGCCATCAACAGCAGA 240
 Db 1075 GGTCAACCTCGGAAACAGTGTGATGTCACCCATGCTGGTGGTGGCCATCAACAGCAGA 1134
 QY 241 TCCTAGTGGCCCCACGACCATCCATGCCAAGTTCACCGACTGGAAGGGTACCTCATGA 300
 Db 1135 TCCTAGTGGCCCCACGACCATCCATGCCAAGTTCACCGACTGGAAGGGTACCTCATGA 1194
 QY 301 AACGGCTGTGGGCTCCAGGACGCTTCCCGTGGATTTCCACATCAAGATGGTGGAGACA 360
 Db 1195 AACGGCTGTGGGCTCCAGGACGCTTCCCGTGGATTTCCACATCAAGATGGTGGAGACA 1254
 QY 361 TGAAGTACCCTTTAGCAGGCGCATGCGGTGGAAGTGTGGACAAGTCCACAGGTGCAC 420
 Db 1255 TGAAGTACCCTTTAGCAGGCGCATGCGGTGGAAGTGTGGACAAGTCCACAGGTGCAC 1314
 QY 421 GCATCGCATGGCTGTGGTGGACACAGTATCGGGGGTGCCTACGGCTCCTTACGAGG 480
 Db 1315 GCATCGCATGGCTGTGGTGGACACAGTATCGGGGGTGCCTACGGCTCCTTACGAGG 1374
 QY 481 ATGGTGACAGTGACGACGACTTCTGTGTCCACATGTGGAGCCCGCTGATCCACCCAGTGG 540
 Db 1375 ATGGTGACAGTGACGACGACTTCTGTGTCCACATGTGGAGCCCGCTGATCCACCCAGTGG 1434
 QY 541 GTTGGTCAACAGTGTGGGCGCAGGATCAAGATGTCAAGAGGGGAGATGACATGGCCCC 600
 Db 1435 GTTGGTCAACAGTGTGGGCGCAGGATCAAGATGTCAAGAGGGGAGATGACATGGCCCC 1494
 QY 601 ATCACCCCACTTCCGGAAGATCTACTGTGTGCGCTTCTTACCTCTTCAAGAAGGTAC 660
 Db 1495 ATCACCCCACTTCCGGAAGATCTACTGTGTGCGCTTCTTACCTCTTCAAGAAGGTAC 1554
 QY 661 GAGCAGTCTACACAGAAGCGGTGGTGGTGGTGGAGAGGATGAAGTGGAGGCGCATGGACC 720
 Db 1555 GAGCAGTCTACACAGAAGCGGTGGTGGTGGTGGAGAGGATGAAGTGGAGGCGCATGGACC 1614
 QY 721 CCTGAATCTGGGACACATCTCGTGGCACTGTCTGTAAAGTTCCTCTGGATGATAC 780
 Db 1615 CCTGAATCTGGGACACATCTCGTGGCACTGTCTGTAAAGTTCCTCTGGATGATAC 1674
 QY 781 TGATGATCTGTGGACGCGGGGGCGCTCCACAGATGGCTTGGACTGGTTCCTACCATG 840
 Db 1675 TGATGATCTGTGGACGCGGGGGCGCTCCACAGATGGCTTGGACTGGTTCCTACCATG 1734

841 CTTCTTCCCAAGCCATCTTCCCGGCCACCTTCTGTGTCAGAGAATGACATTGAGCTCACAC 900
1735 CTTCTTCCCAAGCCATCTTCCCGGCCACCTTCTGTGTCAGAGAATGACATTGAGCTCACAC 1794
901 CGCCAAAAGGTTATGAGGCACAGACTTTCACCTGGGAGGAGTACTTGGAGAGACCAAGT 960
1795 CGCCAAAAGGTTATGAGGCACAGACTTTCACCTGGGAGGAGTACTTGGAGAGACCAAGT 1854
961 CGAAAGCCGCTCCATCGAGACTCTTTAACAATGATTGCCAAACCACTGCTTCAAGGTGG 1020
1855 CGAAAGCCGCTCCATCGAGACTCTTTAACAATGATTGCCAAACCACTGCTTCAAGGTGG 1914
1021 GCATGAAGCTGGAGCGCTGAGCTGATGAGAGCCCGCTCATCTGTGTGGCCACGGTGA 1080
1915 GCATGAAGCTGGAGCGCTGAGCTGATGAGAGCCCGCTCATCTGTGTGGCCACGGTGA 1974
1081 AACGAGTGTGTCATCGGCTCTCTCAGCATCCACTTTGACGGCTGGGACAGGAGTACGACC 1140
1975 AACGAGTGTGTCATCGGCTCTCTCAGCATCCACTTTGACGGCTGGGACAGGAGTACGACC 2034
1141 AGTGGGTGAGTCCGAGTCCCGAGACATCTACCCCGTGGCTGTGTGAGCTCACCGGCT 1200
2035 AGTGGGTGAGTCCGAGTCCCGAGACATCTACCCCGTGGCTGTGTGAGCTCACCGGCT 2094
1201 ACCAGCTCCAGGCTCTGTGGCGCAGAACCGGCCACACCGCTGGAAGGCCAAGAGGCCA 1260
2095 ACCAGCTCCAGGCTCTGTGGCGCAGAACCGGCCACACCGCTGGAAGGCCAAGAGGCCA 2154
1261 CAAAGAAAGAAAGAAACAGTTTGGGAAGAAAGAAAGAAATCCCGCCACTAAGAGCC 1320
2155 CAAAGAAAGAAAGAAACAGTTTGGGAAGAAAGAAAGAAATCCCGCCACTAAGAGCC 2214
1321 GACCCCTCAGACAGAGGGCTCCAGAGAGCCCTCTCGAGAGCAGACCCCTCAGGGTCCAGGA 1380
2215 GACCCCTCAGACAGAGGGCTCCAGAGAGCCCTCTCGAGAGCAGACCCCTCAGGGTCCAGGA 2274
1381 AGATCTCGTCCGAGCTGTCTCTCGCGAGATCAITGCTGTGCGTGTGAAGAAAGAGCATC 1440
2275 AGATCTCGTCCGAGCTGTCTCTCGCGAGATCAITGCTGTGCGTGTGAAGAAAGAGCATC 2334
1441 TAGAGTGGCTCGCCGACAGGCTTCAAGTCCAGAGTGCCTGTCTCCGTCGAGAAC 1500
2335 TAGAGTGGCTCGCCGACAGGCTTCAAGTCCAGAGTGCCTGTCTCCGTCGAGAAC 2394
1501 TCAAGCAGAAACAGACGACTGAGCTTCTCGCTCCAGCCTGGCTTCTAGCTGGAAGCC 1560
2395 TCAAGCAGAAACAGACGACTGAGCTTCTCGCTCCAGCCTGGCTTCTAGCTGGAAGCC 2454
1561 AGCCAGCGTTTCTTACACACACCACTGCTCCACTGACTTTGGCTTGGAGACTGA 1620
2455 AGCCAGCGTTTCTTACACACCACTGCTCCACTGACTTTGGCTTGGAGACTGA 2514
1621 TCCTCTCTGTAAATTCCTCCCGCTGTGTGAAGGCTGGACGCTGGAGGACTGCTGGG 1680
2515 TCCTCTCTGTAAATTCCTCCCGCTGTGTGAAGGCTGGACGCTGGAGGACTGCTGGG 2574
1681 GTCTCTGGACCCCGCTTGTCTTCTCGCTCCCTGTGGAAGGCTATATGACGGGC 1740
2575 GTCTCTGGACCCCGCTTGTCTTCTCGCTCCCTGTGGAAGGCTATATGACGGGC 2634
1741 CGCCTGAGGCCCCAGACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCTCAAGCTG 1800
2635 CGCCTGAGGCCCCAGACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCTCAAGCTG 2694
1801 GAACCTAGTCTGCTCTTCTTCTTAAGATGGCTTCCCGCCAGCCCGCCAGGCTCTCAG 1860
2695 GAACCTAGTCTGCTCTTCTTCTTAAGATGGCTTCCCGCCAGCCCGCCAGGCTCTCAG 2754
1861 TTGCCAGGATGGGCGCCACCACTGTACACTGTGGATACAGACAGTGAACCTGTCTG 1920
2755 TTGCCAGGATGGGCGCCACCACTGTACACTGTGGATACAGACAGTGAACCTGTCTG 2814

1921 CCTGAACGAGTCATGTAATTAAGTCTTAGACAGCTCTCTGAGCAGGATAAGGTCCCT 1980
2815 CCTGAACGAGTCATGTAATTAAGTCTTAGACAGCTCTCTGAGCAGGATAAGGTCCCT 2874
1981 GACAGTGAAGTGTGTGTGGGGGAGCCTCTGCTCAAAAATTCACCAAGCAGATGCCT 2040
2875 GACAGTGAAGTGTGTGTGGGGGAGCCTCTGCTCAAAAATTCACCAAGCAGATGCCT 2934
2041 CTCAGCCTCATGTCTTGGTCTCTGCTCTCTAGCTCCCGAGGATGTTGGGGACCCAG 2100
2935 CTCAGCCTCATGTCTTGGTCTCTGCTCTCTAGCTCCCGAGGATGTTGGGGACCCAG 2994
2101 CTTGTCTCGGAGCTTAAGAGCAGTGACCAAGATGTGGATTTTGGCGAAGCTGTGTGTGG 2160
2995 CTTGTCTCGGAGCTTAAGAGCAGTGACCAAGATGTGGATTTTGGCGAAGCTGTGTGTGG 3054
2161 CTTGAGCTGCTTCTGCTGTGTGTGTGAGGACTGACTCCCATTTCTAAAGGAAATGCCCC 2220
3055 CTTGAGCTGCTTCTGCTGTGTGTGTGAGGACTGACTCCCATTTCTAAAGGAAATGCCCC 3114
2221 GGGGAGCAGATTTGGGAGAGATGGGCTGTGATGTGCACTTTGGCTCTGTCTACCTGTCTCT 2280
3115 GGGGAGCAGATTTGGGAGAGATGGGCTGTGATGTGCACTTTGGCTCTGTCTACCTGTCTCT 3174
2281 GAAGCCCGCTCAAAAATAATTCATCCAAGATTCCTTTGTAGTTAAAGGTCAGTTCTGA 2340
3175 GAAGCCCGCTCAAAAATAATTCATCCAAGATTCCTTTGTAGTTAAAGGTCAGTTCTGA 3234
2341 CTGAGCCTCTAGAGAGCTGGGCTTGTATGTTCTTTGGCTTTTGTCTACCTAAATG 2400
3235 CTGAGCCTCTAGAGAGCTGGGCTTGTATGTTCTTTGGCTTTTGTCTACCTAAATG 3294
2401 AAGAAACCATGCTCGAGGGGCGTGAACACAGAACCTCTCAAGACAGGATGACAGCT 2460
3295 AAGAAACCATGCTCGAGGGGCGTGAACACAGAACCTCTCAAGACAGGATGACAGCT 3354
2461 GGAGGACACATCTAGCTGCAATTCACCTCTAGGCTCCCGAGACTCTGTGTGTGAGA 2520
3355 GGAGGACACATCTAGCTGCAATTCACCTCTAGGCTCCCGAGACTCTGTGTGTGAGA 3414
2521 AATTAAACCCCTGCTTGTGTGAG 2544
3415 AATTAAACCCCTGCTTGTGTGAG 3438

RESULT 4

ABZ71320
ID ABZ71320 standard; cDNA; 3530 BP.

AC ABZ71320;

XX AC
DT 04-APR-2003 (first entry)

XX Secreted protein-encoding gene 131 cDNA clone HOBJ82, SEQ ID NO:141.

XX Human; secreted protein; digestive disorder; gastrointestinal disorder;
XX mouth; esophagus; stomach; small intestine; large intestine; liver;
XX biliary tract; pancreas; cancer; tumour; hyperproliferative disorder;
XX immune disorder; inflammation; infection; wound healing; drug screening;
XX chromosome identification; chromosome mapping; cytostatic;
XX anti-inflammatory; immunosuppressive; vulnery; gene therapy; gene; ss.

OS Homo sapiens.

XX WO200276488-A1.

PN 03-OCT-2002.

XX 19-MAR-2002; 2002WO-US08276.

XX 21-MAR-2001; 2001US-277340P.

PR 19-JUL-2001; 2001US-306171P.

PR 13-NOV-2001; 2001US-331287P.

XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX DR WPI; 2003-029900/02.
XX DR P-PSDB; ABR00141.
XX PT New human secreted proteins and nucleic acids, useful for detecting,
XX PT preventing, diagnosing, prognosticating, treating and/or ameliorating
XX PT e.g. gastrointestinal diseases and disorders, or cancers -
XX PS Claim 21; Page 834-835; 1216pp; English.
XX PS ABZ71190-ABZ71478 represent cDNAs corresponding to 178 human secreted
XX CC protein genes, and ABR00011-ABR00299 represent the proteins they encode.
XX CC ABZ71479-ABZ71540 represent human secreted protein genomic fragments. The
XX CC invention also encompasses antibodies specific for the secreted proteins,
XX CC the use of the secreted proteins in drug screening, and recombinant
XX CC vectors and host cells comprising a nucleic acid of the invention. The
XX CC secreted proteins, nucleic acids encoding them, antibodies or antibody
XX CC fragments specific for the secreted proteins, and modulators of protein
XX CC activity are useful for the diagnosing, treating, ameliorating or preventing
XX CC digestive disorders. Such conditions include disorders of the mouth,
XX CC oesophagus, stomach, and include cancers of these organs and tissues. The
XX CC secreted proteins and their nucleic acids may also be used in the
XX CC treatment of immune disorders, inflammation, infection,
XX CC hyperproliferative disorders, and to promote wound healing. Nucleic acids
XX CC of the invention may be used for chromosome identification, chromosome
XX CC mapping, in gene therapy, for identifying individuals from minute
XX CC biological samples, as hybridisation probes, and as molecular weight
XX CC markers. The present sequence represents a human secreted protein-
XX CC encoding cDNA clone of the invention.
XX SQ Sequence 3530 BP; 815 A; 945 C; 972 G; 790 T; 8 other;

Query Match 99.6%; Score 2544; DB 25; Length 3530;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGTCCCACTCTATACAGAGTGGAGGATGATGATGAAGAGGATGAGGTGAGGTGC 60
DB 895 AGCAGTCCCACTCTATACAGAGTGGAGGATGATGATGAAGAGGATGAGGTGAGGTGC 954
QY 61 TCAACAGTGTGCTGTGCTCCCAAGCCGGGTGTACTGGATCGCTCTGTCTATCCAGACAG 120
DB 955 TCAACAGTGTGCTGTGCTCCCAAGCCGGGTGTACTGGATCGCTCTGTCTATCCAGACAG 1014
QY 121 CAGGATATCGGGTGTGCTTTCGGTATGAAGCTTTGAAATGACGCCAGCATGACTTCT 180
DB 1015 CAGGATATCGGGTGTGCTTTCGGTATGAAGCTTTGAAATGACGCCAGCATGACTTCT 1074
QY 181 GGTCAAGCTGGGAACAGTGTGATGCCACCCATTTGGTGTGGTGTGCCATCAACAGCAAGA 240
DB 1075 GGTCAAGCTGGGAACAGTGTGATGCCACCCATTTGGTGTGGTGTGCCATCAACAGCAAGA 1134
QY 241 TCCTAGTCCCTCCAGGACCATTCATGCCAAGTTTCCAGCTGGAAGGGTACCTCATGA 300
DB 1135 TCCTAGTCCCTCCAGGACCATTCATGCCAAGTTTCCAGCTGGAAGGGTACCTCATGA 1194
QY 301 AACGGCTGGTGGCTCCAGGACGCTTCCGTTGGATTTCACATCAAGATGGTGAGAGCA 360
DB 1195 AACGGCTGGTGGCTCCAGGACGCTTCCGTTGGATTTCACATCAAGATGGTGAGAGCA 1254
QY 361 TGAAGTACCCCTTTAGGCAGGGCATGCGCTGGAGTGGTGGACAAAGTCCAGGTGTAC 420
DB 1255 TGAAGTACCCCTTTAGGCAGGGCATGCGCTGGAGTGGTGGACAAAGTCCAGGTGTAC 1314
QY 421 GCATTCGATCGCTGTGTGTGACACAGTAATTCGGGGTGTGGCTTACGGCTCTCTACGAGG 480
DB 1315 GCATTCGATCGCTGTGTGTGACACAGTAATTCGGGGTGTGGCTTACGGCTCTCTACGAGG 1374

QY 481 ATGTTGACAGTGCAGCAGCTTCTGTGTGCACATGTGGAGCCCTGTATCCACCCAGTGG 540
DB 1375 ATGTTGACAGTGCAGCAGCTTCTGTGTGCACATGTGGAGCCCTGTATCCACCCAGTGG 1434
QY 541 GTTGGTCAACAGCTGTGGCCACGCGATCAAGATGTTCAGAGAGCGGAGATGATGGCCC 600
DB 1435 GTTGGTCAACAGCTGTGGCCACGCGATCAAGATGTTCAGAGAGCGGAGATGATGGCCC 1494
QY 601 ATCAACCCACCTTTCGGAAAGATCTACTGTGATGCCCTTCTTACCTCTTCAAGAGGTAC 660
DB 1495 ATCAACCCACCTTTCGGAAAGATCTACTGTGATGCCCTTCTTACCTCTTCAAGAGGTAC 1554
QY 661 GAGCAGTCTACACAGAGCGGCTGTGTTGAGAGAGGATGAAGCTGGAGGCCATTTGACC 720
DB 1555 GAGCAGTCTACACAGAGCGGCTGTGTTGAGAGAGGATGAAGCTGGAGGCCATTTGACC 1614
QY 721 CCTGATCTGGGCAACATCTCGGTGGCAACTGTCTGTAAAGGTTCCTCTGGATGGATACC 780
DB 1615 CCTGATCTGGGCAACATCTCGGTGGCAACTGTCTGTAAAGGTTCCTCTGGATGGATACC 1674
QY 781 TGATGATCTGTGTGAGAGCGGGGCCCTCCACAGATGGCTTGGACTGGTCTGTCTACATG 840
DB 1675 TGATGATCTGTGTGAGAGCGGGGCCCTCCACAGATGGCTTGGACTGGTCTGTCTACATG 1734
QY 841 CCTCTTCCACAGCGCATCTTCCCGGCCACCTTCTGTCAAGAGATGACATTTGAGCTCACAC 900
DB 1735 CCTCTTCCACAGCGCATCTTCCCGGCCACCTTCTGTCAAGAGATGACATTTGAGCTCACAC 1794
QY 901 CGCCAAAAGGTTATGAGGCACAGACTTTTCAACTGGGAGAACTACTTTGGAGAGACCAAGT 960
DB 1795 CGCCAAAAGGTTATGAGGCACAGACTTTTCAACTGGGAGAACTACTTTGGAGAGACCAAGT 1854
QY 961 CGAAGCGCTCCATCGAGACTCTTTAAATGATGGATGCCAAACCATTTGAGCTTCAAGTGG 1020
DB 1855 CGAAGCGCTCCATCGAGACTCTTTAAATGATGGATGCCAAACCATTTGAGCTTCAAGTGG 1914
QY 1021 GCATGAAGCTGGAGCGGTGGACCTGATGGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1080
DB 1915 GCATGAAGCTGGAGCGGTGGACCTGATGGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1974
QY 1081 AACGAGTGTGATCGGCTCTCAGATCCATTTGAAGGCTGGGACAGCGAGTACGACC 1140
DB 1975 AACGAGTGTGATCGGCTCTCAGATCCATTTGAAGGCTGGGACAGCGAGTACGACC 2034
QY 1141 AGTGGGTGACTGCGAGTCCCGCAGACATCTACCCGCTGGGCTGTGTGAGCTCACCGGCT 1200
DB 2035 AGTGGGTGACTGCGAGTCCCGCAGACATCTACCCGCTGGGCTGTGTGAGCTCACCGGCT 2094
QY 1201 ACCAGCTCCAGCTCTGTGTGGCCGAGAACCCGGCCACACCCGCTGAAGGCCAAGAGGCCA 1260
DB 2095 ACCAGCTCCAGCTCTGTGTGGCCGAGAACCCGGCCACACCCGCTGAAGGCCAAGAGGCCA 2154
QY 1261 CAAAGAGAAAGAAACAGATTTGGGAGAAAGAAAGAAAGAAATCCCGCCACTAAGAGCC 1320
DB 2155 CAAAGAGAAAGAAACAGATTTGGGAGAAAGAAAGAAAGAAATCCCGCCACTAAGAGCC 2214
QY 1321 GACCCCTCAGACAGGGGTCCAAAGAGCCCTCTCTGGAGAGCAGCCCTCAGGCTGCCAGGA 1380
DB 2215 GACCCCTCAGACAGGGGTCCAAAGAGCCCTCTCTGGAGAGCAGCCCTCAGGCTGCCAGGA 2274
QY 1381 AGATCTGTGCGAGCTGTCTCTGGCGAGATCATTTGTGTGTGTGTGTGTGTGTGTGTGTGT 1440
DB 2275 AGATCTGTGCGAGCTGTCTCTGGCGAGATCATTTGTGTGTGTGTGTGTGTGTGTGTGTGT 2334
QY 1441 TAGAGTGGCTTCGCGCCGCAAGAGTTCAGGTCCAGAGCTGTCTCTCTCGCTCGAGAGCA 1500
DB 2335 TAGAGTGGCTTCGCGCCGCAAGAGTTCAGGTCCAGAGCTGTCTCTCTCGCTCGAGAGCA 2394
QY 1501 TCAAGCAGAGAAACAGACAGCTGAGCTTCTCTGCTCCAGCTTGGCTTCTAGCTGGAAGCC 1560
DB 2395 TCAAGCAGAGAAACAGACAGCTGAGCTTCTCTGCTCCAGCTTGGCTTCTAGCTGGAAGCC 2454
QY 1561 AGCCACAGCGTCTTCTCTACACACACCATGCCTCCATGCTTGTGTGTGTGTGTGTGTGTGTGT 1620

Db	2455	AGCCACGCTTCTCTACACACCACTGCCTCCACCTGACTTTGGCTTGGAGACTGA	2514
Qy	1621	TCCTCTCTGTGTAAATTCCTCCGGTCTGTGAAGCTGGACGCTGAGACCTGTCTGG	1680
Db	2515	TCCTCTCTGTGTAAATTCCTCCGGTCTGTGAAGCTGGACGCTGAGACCTGTCTGG	2574
Qy	1681	GTCTCTCTGGACCGCTGTCTTCCTGCGCCCTCCCTGTGGAAAGTCTATATGACGGGC	1740
Db	2575	GTCTCTCTGGACCGCTGTCTTCCTGCGCCCTCCCTGTGGAAAGTCTATATGACGGGC	2634
Qy	1741	CGCCTGAGGCCCCAGAACTCGTCTGTGAACCACTTTTCAGCCAGAGTTCCCAAGCTG	1800
Db	2635	CGCCTGAGGCCCCAGAACTCGTCTGTGAACCACTTTTCAGCCAGAGTTCCCAAGCTG	2694
Qy	1801	GAAGCTAGCTGCTGCTCTCTCTTAAGATGGCTCCCTCCCGACCCGCGACGGCCCTCAG	1860
Db	2695	GAAGCTAGCTGCTGCTCTCTCTTAAGATGGCTCCCTCCCGACCCGCGACGGCCCTCAG	2754
Qy	1861	TTGCCAGGGATGGGCCACCACTGTACACTGTGTGGAAACAAGACAGTGAACCTGTCTG	1920
Db	2755	TTGCCAGGGATGGGCCACCACTGTACACTGTGTGGAAACAAGACAGTGAACCTGTCTG	2814
Qy	1921	CCTGAACGAGTCATGAATAAAGTTCTAGACAGCTCTCTGACGAGTAAGGTCCCT	1980
Db	2815	CCTGAACGAGTCATGAATAAAGTTCTAGACAGCTCTCTGACGAGTAAGGTCCCT	2874
Qy	1981	GACAGTGAATTGTGTGGGGCAGCCTCTGCTCAAAAATTCACACAGCAAAATGCCT	2040
Db	2875	GACAGTGAATTGTGTGGGGCAGCCTCTGCTCAAAAATTCACACAGCAAAATGCCT	2934
Qy	2041	CTCAGCCTCATGTGTTGGTCTCTGCTCTCTAGCTCCCGAGGATGTTGGGGACCCAG	2100
Db	2935	CTCAGCCTCATGTGTTGGTCTCTGCTCTCTAGCTCCCGAGGATGTTGGGGACCCAG	2994
Qy	2101	CTTGCTCTGGCAGCTAAGAAGCAGTACACAGGATGTGGATTTTGGCGACCTGTGTGGTGG	2160
Db	2995	CTTGCTCTGGCAGCTAAGAAGCAGTACACAGGATGTGGATTTTGGCGACCTGTGTGGTGG	3054
Qy	2161	CCTTGAGCTGCTTTCTGTGTTTGTGAGGACTGACTCCCATTTCTAAAGAAATGCCCC	2220
Db	3055	CCTTGAGCTGCTTTCTGTGTTTGTGAGGACTGACTCCCATTTCTAAAGAAATGCCCC	3114
Qy	2221	GGGGAGGACATTTGGGAGGAGATGGCTGAGTGTGCATTTGGCTCTGCTACTGCTCTCT	2280
Db	3115	GGGGAGGACATTTGGGAGGAGATGGCTGAGTGTGCATTTGGCTCTGCTACTGCTCTCT	3174
Qy	2281	GAAGCCCGTAAATAAATCATCCAGATTCCTTTGTAGTTAAAGGCTCCAGTTCTGA	2340
Db	3175	GAAGCCCGTAAATAAATCATCCAGATTCCTTTGTAGTTAAAGGCTCCAGTTCTGA	3234
Qy	2341	CTGGAGCCTCTAGAGAGCTGGGCTGTATGTTCTTTTGGCTTTTGTCTCTACTAAATG	2400
Db	3235	CTGGAGCCTCTAGAGAGCTGGGCTGTATGTTCTTTTGGCTTTTGTCTCTACTAAATG	3294
Qy	2401	AAGAACCATTGCTGAGGGGCCGTGAACACAGAACCTCAGACAAAGGATGACAGCT	2460
Db	3295	AAGAACCATTGCTGAGGGGCCGTGAACACAGAACCTCAGACAAAGGATGACAGCT	3354
Qy	2461	GGAGGACACATCTAGCTGCCATTGCAACCTCACTGGGCTCCCGACAGCTCTGTGTGGAGA	2520
Db	3355	GGAGGACACATCTAGCTGCCATTGCAACCTCACTGGGCTCCCGACAGCTCTGTGTGGAGA	3414
Qy	2521	AATTAACCCCTGCTTGTCTGAG	2544
Db	3415	AATTAACCCCTGCTTGTCTGAG	3438

RESULT 5
ABL89857
ID ABL89857 standard; CDNA; 2663 bp.
XX
AC ABL89857;

XX	24-MAY-2002 (first entry)	
DT	Human polynucleotide SEQ ID NO 419.	
XX		
DE		
XX		
XX	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;	
KW	antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;	
KW	vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;	
KW	cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;	
KW	neurological disease; infection; human; secreted protein; gene; ss.	
XX		
OS	Homo sapiens.	
XX		
XX	WO2001/90304-A2.	
PN		
XX	29-NOV-2001.	
PD		
XX	18-MAY-2001; 2001WO-US16450.	
XX		
PR	19-MAY-2000; 2000US-205515P.	
XX		
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA		
XX	Birse CE, Rosen CA;	
PI		
XX	WPI; 2002-122018/16.	
DR	P-PSDB; ABB9448.	
DR		
XX	Novel 1405 isolated polypeptides, useful for diagnosis, treatment and	
PT	prevention of neural, immune system, muscular, reproductive, and	
PT	gastrointestinal, pulmonary, cardiovascular, renal and proliferative	
PT	disorders -	
XX		
XX	Claim 4; SEQ ID NO 419; 2081pp + Sequence Listing; English.	
PS		
XX	The invention relates to novel genes (AB189449-ABL90853) and proteins	
CC	(ABB89040-ABB9044) useful for preventing, treating or ameliorating	
CC	medical conditions e.g. by protein or gene therapy. The genes are	
CC	isolated from a range of human tissues disclosed in the specification.	
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful	
CC	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast	
CC	and ovarian cancer and other cancers of the adrenal gland, bone, bone	
CC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;	
CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune	
CC	haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's	
CC	disease, multiple sclerosis, rheumatoid arthritis and ulcerative	
CC	colitis; (c) cardiovascular disorders such as myocardial ischaemias;	
CC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and	
CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal	
CC	and parasitic infections.	
CC	Note: The sequence data for this patent did not form part of the	
CC	printed specification, but was obtained in electronic format directly	
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.	
XX		
SQ	Sequence 2663 BP; 589 A; 731 C; 751 G; 584 T; 8 other;	
	Query Match 98.4%; Score 2514.4; DB 24; Length 2663;	
	Best Local Similarity 99.7%; Pred. No. 0;	
	Matches 2537; Conservative 1; Mismatches 5; Indels 2; Gaps 2;	
QY	1 AGCAGTCCACTCTATGACCAAGTGGAGGATGATGAAGGGATGAAGTGGAGGTGC 60	
DB	72 AGCAGTCCACTCTATGACCAAGTGGAGGATGATGAAGGGATGAAGTGGAGGTGC 131	
QY	61 TCACACAGTGATCGTGTGCTCCCGACCGGGTACTGGATCGCTCTGTTCATCCAGACAG 120	
DB	132 TCACACAGTGATCGTGTGCTCCCGACCGGGTACTGGATCGCTCTGTTCATCCAGACAG 191	
QY	121 CAGGGTATCGGGTGTGCTTCGGTATGAGGCTTTGAAATGACCCAGCCATGACTTCT 180	
DB	192 CAGGGTATCGGGTGTGCTTCGGTATGAGGCTTTGAAATGACCCAGCCATGACTTCT 251	
QY	181 GGTGCACCTCGGAACAGTGGATGTCCACCCCATTTGGTGTGTGCTTCATCCAGACAGA 240	

Db 252 GGTGCACTGGGAAACAGTGGATGTCCACCCCATTTGGCTGGTGCATCAACAGCAAGA 311
Qy 241 TCTTAGTGGCCCCCAGGACCATCATGCCAAGTTCACCGACTGGAAGGGCTACTCATGA 300
Db 312 TCTTAGTGGCCCCCAGGACCATCATGCCAAGTTCACCGACTGGAAGGGCTACTCATGA 371
Qy 301 AACGGCTGGTGGGCTCCAGAGAGCTTCCCGTGAGATTCACATCAAGATGGTGGAGAGCA 360
Db 372 AACGGCTGGTGGGCTCCAGAGAGCTTCCCGTGAGATTCACATCAAGATGGTGGAGAGCA 431
Qy 361 TGAAGTACCCCTTTAGSCAGGGCATGCGGTGGAAGTGGTGAAGTCCCAAGTGTCTAC 420
Db 432 TGAAGTACCCCTTTAGSCAGGGCATGCGGTGGAAGTGGTGAAGTCCCAAGTGTCTAC 491
Qy 421 GCACCTGCATGGTGGTGGACACAGTAAATCGGGGGTGCCTACCGCTTCCTCTACGAGG 480
Db 492 GCACCTGCATGGTGGTGGACACAGTAAATCGGGGGTGCCTACCGCTTCCTCTACGAGG 551
Qy 481 ATGGTGACAGTGACGACGACTTCTGGTGGCCACATGTGGAGGCCCTGATCCACCCAGTGG 540
Db 552 ATGGTGACAGTGACGACGACTTCTGGTGGCCACATGTGGAGGCCCTGATCCACCCAGTGG 611
Qy 541 GTTGGTCAAGCATGTGGGCCACCGCATCAAGATGTCAAGAGGGCAAGTGAATGGCCC 600
Db 612 GTTGGTCAAGCATGTGGGCCACCGCATCAAGATGTCAAGAGGGCAAGTGAATGGCCC 671
Qy 601 ATCACCCCACTTCCGGAAGATCTACTGTGATGCCGTTCCTTACCTTTCAGAGAGGTAC 660
Db 672 ATCACCCCACTTCCGGAAGATCTACTGTGATGCCGTTCCTTACCTTTCAGAGAGGTAC 731
Qy 661 GAGCAGTCTACAGAAAGCGGTGGTTTGAAGAGGGATGAAGCTGGAGGCCATTTGACC 720
Db 732 GAGCAGTCTACAGAAAGCGGTGGTTTGAAGAGGGATGAAGCTGGAGGCCATTTGACC 791
Qy 721 CCTGTAACTGGGCACATCTGGTGGCAACTGTCTGAAGTTCTCTGGATGATAC 780
Db 792 CCTGTAACTGGGCACATCTGGTGGCAACTGTCTGAAGTTCTCTGGATGATAC 851
Qy 781 TGATGATCTGTGGACGGGGGGCCCTCCACAGATGGCTTGACACTGGTTCTGTACCATG 840
Db 852 TGATGATCTGTGGACGGGGGGCCCTCCACAGATGGCTTGACACTGGTTCTGTACCATG 911
Qy 841 CCTCTTCCACGCAATCTTCCGGGCCACTTCTGTCAAGAAATGACATTTGAGTCTCAC 900
Db 912 CCTCTTCCACGCAATCTTCCGGGCCACTTCTGTCAAGAAATGACATTTGAGTCTCAC 971
Qy 901 CGCCAAAGGTTATGAGGCACAGACTTTCACCTGGGAGAACTACTTGGAGAGACCAAGT 960
Db 972 CGCCAAAGGTTATGAGGCACAGACTTTCACCTGGGAGAACTACTTGGAGAGACCAAGT 1031
Qy 961 CGAAAGCCGCTCCATCGAGACTCTTTAAATGGAATGGCCCAACCATGGCTTCAAGGTGG 1020
Db 1032 CGAAAGCCGCTCCATCGAGACTCTTTAAATGGAATGGCCCAACCATGGCTTCAAGGTGG 1090
Qy 1021 GCATGAGCTGAGGCGCTGGACCTGATGAGAGCCCGGCTCATCTGTGGCCACGGTGA 1080
Db 1091 GCATGAGCTGAGGCGCTGGACCTGATGAGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1150
Qy 1081 AACGAGTGGTGCATCGGCTCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACGACC 1140
Db 1151 AACGAGTGGTGCATCGGCTCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACGACC 1210
Qy 1141 AGTGGTGGACTGGAGTGGCCAGACATCTACCCCGT-CGGCTGGTGGAGCTCACCGGC 1199
Db 1211 AGTGGTGGACTGGAGTGGCCAGACATCTACCCCGTTCGCGTGGTGGAGCTCACCGGC 1270
Qy 1200 TACAGCTCCAGCTTCTGTGGCCGAGAACCGGCCACACCGCTGAAGGCCCAAGAGGCC 1259
Db 1271 TACAGCTCCAGCTTCTGTGGCCGAGAACCGGCCACACCGCTGAAGGCCCAAGAGGCC 1330
Qy 1260 ACAAGAGAGAAAGAAACAGTTTGGGAGAAAGAAAGATCCCGCCCACTAAGACG 1319

Db 1331 ACAAGAGAGAAAGAAACAGTTTGGGAGAAAGAAAGAAATCCCGCCCACTAAGACG 1390
Qy 1320 CGACCCCTCAGCAGGGGTCCAAAGAGCCCTGTGGAGAGCAGCCCTCAGGGTGGCCAGG 1379
Db 1391 CGACCCCTCAGCAGGGGTCCAAAGAGCCCTGTGGAGAGCAGCCCTCAGGGTGGCCAGG 1450
Qy 1380 AAGATCTCTGTGGAGCTTTCCTGGGAGATCATTTCTGTGCGTGTGAAGAGAGAGCAT 1439
Db 1451 AAGATCTCTGTGGAGCTTTCCTGGGAGATCATTTCTGTGCGTGTGAAGAGAGAGCAT 1510
Qy 1440 CTAGACGTGGGCTCGCCGACAGAGCTTCAAGTCCAGAGCTGCGCTCTAGCTGGAGAC 1499
Db 1511 CTAGACGTGGGCTCGCCGACAGAGCTTCAAGTCCAGAGCTGCGCTCTAGCTGGAGAC 1570
Qy 1500 ATCAAGCAGGAGAAACAGACGACTGAGCCCTTCTGCTCCAGCTGGCTTCTAGCTGGAGC 1559
Db 1571 ATCAAGCAGGAGAAACAGACGACTGAGCCCTTCTGCTCCAGCTGGCTTCTAGCTGGAGC 1630
Qy 1560 CAGCCAGAGCTTCTCTACACACACACCATGCTCCACCTGACTTTGGCTGGAGACTG 1619
Db 1631 CAGCCAGAGCTTCTCTACACACACACCATGCTCCACCTGACTTTGGCTGGAGACTG 1690
Qy 1620 ATCTCTCTGTGTAATTTCTGCCGGTGTGTGAAGCTGAGCGGTGGAGACCTGTGG 1679
Db 1691 ATCTCTCTGTGTAATTTCTGCCGGTGTGTGAAGCTGAGCGGTGGAGACCTGTGG 1750
Qy 1680 GGTCTCTCTGGACCCGCTGTGTGCTGCTCCCTCCCTGTGGAAAGGTCTATATGACGG 1739
Db 1751 GGTCTCTCTGGACCCGCTGTGTGCTGCTCCCTCCCTGTGGAAAGGTCTATATGACGG 1810
Qy 1740 CCGCTCAGGCCCCCAGAACTCGTGTGAACACACTTTTCCAGCCAGAGTCCCAAGCT 1799
Db 1811 CCGCTCAGGCCCCCAGAACTCGTGTGAACACACTTTTCCAGCCAGAGTCCCAAGCT 1870
Qy 1800 GBAACGCTAGCTGCTCTTCTTAAAGATGSCCTCCCTCCCGACCCGACCGCCCTCA 1859
Db 1871 GBAACGCTAGCTGCTCTTCTTAAAGATGSCCTCCCTCCCGACCCGACCGCCCTCA 1930
Qy 1860 GTTGCCAGGATGGGCCACCACTGTCACTGTGGAAATACAGACAGTGAATCTGTCT 1919
Db 1931 GTTGCCAGGATGGGCCACCACTGTCACTGTGGAAATACAGACAGTGAATCTGTCT 1990
Qy 1920 GCTGAACGAGTCAATGTAATTAAGTTCTAGACAGCTCTCTGAGCAGGATAAGTCCCC 1979
Db 1991 GCTGAACGAGTCAATGTAATTAAGTTCTAGACAGCTCTCTGAGCAGGATAAGTCCCC 2050
Qy 1980 TGACAGTGAATGTGTGGTGGGGGAGCCCTGCTCCCTCAAAATTCACAGAGAGATGCC 2039
Db 2051 TGACAGTGAATGTGTGGTGGGGGAGCCCTGCTCCCTCAAAATTCACAGAGAGATGCC 2110
Qy 2040 TCTCAGCTCATGTGTGGTTCCTCTCTCTCTAGCTCCCGAGGATGTGGGGACCCA 2099
Db 2111 TCTCAGCTCATGTGTGGTTCCTCTCTCTCTAGCTCCCGAGGATGTGGGGACCCA 2170
Qy 2100 GCTTGTCTGGCAGCTAAGAGCAGTGAACAGATGTGAATTTGGCGACCTGTGTGGTG 2159
Db 2171 GCTTGTCTGGCAGCTAAGAGCAGTGAACAGATGTGAATTTGGCGACCTGTGTGGTG 2230
Qy 2160 GCTTGTGAGCTTCTGTGTGGTGGAGACTGACTCCCATTTCTTAAAGGAAATGCCCC 2219
Db 2231 GCTTGTGAGCTTCTGTGTGGTGGAGACTGACTCCCATTTCTTAAAGGAAATGCCCC 2290
Qy 2220 CGGGAGGACATTTGGAGGAAAGATGSCCTGAGTGTGCTCTTTGGCTCTCTACCTGCTCC 2279
Db 2291 CGGGAGGACATTTGGAGGAAAGATGSCCTGAGTGTGCTCTTTGGCTCTCTACCTGCTCC 2350
Qy 2280 TGAAGCCCGCTTAAATTAATTCATCCAGATTCCTTTAGTAAAGGTCAGTCTG 2339
Db 2351 TGAAGCCCGCTTAAATTAATTCATCCAGATTCCTTTAGTAAAGGTCAGTCTG 2410
Qy 2340 ACTGGAGCTCTAGAGAGCTGGGCTGTGTGTTCTTTTGGCTTTTGTCTTACCTAAAT 2399
Db 2411 ACTGGAGCTCTAGAGAGCTGGGCTGTGTGTTCTTTTGGCTTTTGTCTTACCTAAAT 2470

QY 2400 GAAGAAACCATGCTGGAGGGGCGGTGAACACAGAACCCCTCAGACAAAGATGACAGAGC 2459
 Db 2471 GAAGAAACCATGCTGGAGGGGCGGTGAACACAGAACCCCTCAGACAAAGATGACAGAGC 2530
 QY 2460 TGGAGGACACATCTAGCTGCACTGCAACCTCACTGGGCTCCCGAGACTCTGTGTGAG 2519
 Db 2531 TGGAGGACACATCTAGCTGCACTGCAACCTCACTGGGCTCCCGAGACTCTGTGTGAG 2590
 QY 2520 AATTTAAACCCCTGCTTGTGAG 2544
 Db 2591 AATTTAAACCCCTGCTTGTGAG 2615

RESULT 6

AAI60518/C
 ID AAI60518 standard; cDNA; 2380 BP.
 XX
 AC AAI60518;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 4507.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; SS.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0682191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren P, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue A, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 XX
 DR P-PSDB; AAM41362.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1; SEQ ID NO 4507; 10078pp; English.
 XX
 SS The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 2380 BP; 510 A; 664 C; 670 G; 536 T; 0 other;
 Query Match 92.9%; Score 2374.2; DB 22; Length 2380;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2376; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 176 CTTCTGTGCAACCTGGGAAACAGTGGATGCCACCCATTGGTGTGTGCCATCAACAG 235
 Db 2379 CTTCTGTGCAACCTGGGAAACAGTGGATGCCACCCATTGGTGTGTGCCATCAACAG 2320
 QY 236 CAAGATCCTAGTGGCCCCCAGGACCATCCATGCCAAGTTCACCGACTGGAGGGCTACCT 295
 Db 2319 CAAGATCCTAGTGGCCCCCAGGACCATCCATGCCAAGTTCACCGACTGGAGGGCTACCT 2260
 QY 296 CATGAAACGGCTGGTGGGCTCCAGGACGCTTCCCGTGGATTTCACATCAAGATGGTGA 355
 Db 2259 CATGAAACGGCTGGTGGGCTCCAGGACGCTTCCCGTGGATTTCACATCAAGATGGTGA 2200
 QY 356 GAGCATGAAGTACCCCTTTAGGACGGGATCGGCTGGAGTGGTGGACAGTCCAGGT 415
 Db 2199 GAGCATGAAGTACCCCTTTAGGACGGGATCGGCTGGAGTGGTGGACAGTCCAGGT 2140
 QY 416 GTCACGCACTCGCATGGCTGTGTGTGGACACAGTAATCGGGGTCGCTACGGCTCCTCTA 475
 Db 2139 GTCACGCACTCGCATGGCTGTGTGTGGACACAGTAATCGGGGTCGCTACGGCTCCTCTA 2080
 QY 476 CGAGATGGTGACAGTGACAGCACTTCTGGTGCCACATGTGGAGAGCCCTGATCCACC 535
 Db 2079 CGAGATGGTGACAGTGACAGCACTTCTGGTGCCACATGTGGAGAGCCCTGATCCACC 2020
 QY 536 AGTGGTGTGTGACAGCTGTGGGCCACGCGCATCAAGATGTCAGAGAGGCGAAGTGACAT 595
 Db 2019 AGTGGTGTGTGACAGCTGTGGGCCACGCGCATCAAGATGTCAGAGAGGCGAAGTGACAT 1960
 QY 596 GGCCCATCACCCCACTTCCGGAAGATCTACTGTGATGCCGTTCTTACCTTCAAGAA 655
 Db 1959 GGCCCATCACCCCACTTCCGGAAGATCTACTGTGATGCCGTTCTTACCTTCAAGAA 1900
 QY 656 GGTACGACAGCTCTACACAGAGGCGGTGTGTGTGGAGAGGAGATGAAGCTGGAGGCCAT 715
 Db 1899 GGTACGACAGCTCTACACAGAGGCGGTGTGTGTGGAGAGGAGATGAAGCTGGAGGCCAT 1840
 QY 716 TGACCCCTGAACTCTGGGCAACATCTGCTGGGCAACTGTCTGTAAAGTTCTCTGGATGG 775
 Db 1839 TGACCCCTGAACTCTGGGCAACATCTGCTGGGCAACTGTCTGTAAAGTTCTCTGGATGG 1780
 QY 776 ATACTGATGATCTGTGTGGAGGGGGCCCTCCACAGATGGCTTGGACTGTTCTGCTA 835
 Db 1779 ATACTGATGATCTGTGTGGAGGGGGCCCTCCACAGATGGCTTGGACTGTTCTGCTA 1720
 QY 836 CCATGCTCTTCCCAAGCCATCTTCCCGGCCACCTTCTGTGAGAGAAATGACATTGAGCT 895
 Db 1719 CCATGCTCTTCCCAAGCCATCTTCCCGGCCACCTTCTGTGAGAGAAATGACATTGAGCT 1660
 QY 896 CACACCGCCAAAGGTTATAGGACAGACTTTTCAACTGGGAGAACTACTTGGAGAGAC 955
 Db 1659 CACACCGCCAAAGGTTATAGGACAGACTTTTCAACTGGGAGAACTACTTGGAGAGAC 1600
 QY 956 CAAGTCCGAAGCCGCTCCATCGAGACTCTTTAAATGATGTTGCCAACCACTGGCTTCAA 1015
 Db 1599 CAAGTCCGAAGCCGCTCCATCGAGACTCTTTAAATGATGTTGCCAACCACTGGCTTCAA 1540
 QY 1016 GGTGGGATGAAGCTGGAGGCGGCTGGAGCCCGGCTCATCTGTGTGGCCAC 1075
 Db 1539 GGTGGGATGAAGCTGGAGGCGGCTGGAGCCCGGCTCATCTGTGTGGCCAC 1480
 QY 1076 GGTGAAACGAGTGTGTGATCGCTCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTA 1135

Db 1479 GGTGAACAGTGTGTGATCGCTCCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTA 1420
 QY 1136 CGACAGTGGTGGAGTCCGAGTCCCGAGATATACCCCGTGGCTGGTGTGAGTCCAC 1195
 Db 1419 CGACAGTGGTGGAGTCCGAGTCCCGAGATATACCCCGTGGCTGGTGTGAGTCCAC 1360
 QY 1196 CGGTACAGTCCAGCTCCTGTGGCGGAGAACCGGACACACCGCTGAAGGCCAAAGA 1255
 Db 1359 CGGTACAGTCCAGCTCCTGTGGCGGAGAACCGGACACACCGCTGAAGGCCAAAGA 1300
 QY 1256 GGCACAAAGAAAGAAAGAAAGTGGGAGAAAGAAAGAAAGTCCCGGCCACTAA 1315
 Db 1299 GGCACAAAGAAAGAAAGAAAGTGGGAGAAAGAAAGAAAGTCCCGGCCACTAA 1240
 QY 1316 GACGGACCCCTCAGACAGGGGTCCAGAGCCCTGTGGAGAGACACCCCTCAGGGTGC 1375
 Db 1239 GACGCGACCCCTCAGACAGGGGTCCAGAGCCCTGTGGAGAGACACCCCTCAGGGTGC 1180
 QY 1376 CAGGAAGATCTCGTCGGAGCCTGTTCCTGGGAGATCATTTGCTGTGCTGTGAAGGAAGA 1435
 Db 1179 CAGGAAGATCTCGTCGGAGCCTGTTCCTGGGAGATCATTTGCTGTGCTGTGAAGGAAGA 1120
 QY 1436 GCATCTAGACGTGGCTCGCCGACAAAGCTTCAAGTCCAGAGCTGCTCTCCGCTCGA 1495
 Db 1119 GCATCTAGACGTGGCTCGCCGACAAAGCTTCAAGTCCAGAGCTGCTCTCCGCTCGA 1060
 QY 1496 GAAATCATCAGCAGGAACAGACGACTGAGCTTCTGCTCCAGCTCCAGCTGGCTTCTAGCTGG 1555
 Db 1059 GAAATCATCAGCAGGAACAGACGACTGAGCTTCTGCTCCAGCTGGCTTCTAGCTGG 1000
 QY 1556 AAGCAGCCAGCTGCTTCTTACCAACACCACTGCTCCAGCTGCTTGGCTTGGAG 1615
 Db 999 AAGCAGCCAGCTGCTTCTTACCAACACCACTGCTCCAGCTGCTTGGCTTGGAG 940
 QY 1616 ACTGATCTCTCTGTGTAATTTCTGCGGCTGTGTGAAGCTGGAGCGTGGAGACCTG 1675
 Db 939 ACTGATCTCTCTGTGTAATTTCTGCGGCTGTGTGAAGCTGGAGCGTGGAGACCTG 880
 QY 1676 CTGGGCTCTCTGGGACCGCTGTGCTTCTGCTCCCTGCTGGAAAGTCTATATGA 1735
 Db 879 CTGGGCTCTCTGGGACCGCTGTGCTTCTGCTCCCTGCTGGAAAGTCTATATGA 820
 QY 1736 CGGCGCGCTCAGGCGCCAGAACTGCTGTGAACCACTTTTCCAGCAGAGTTCOCOA 1795
 Db 819 CGGCGCGCTCAGGCGCCAGAACTGCTGTGAACCACTTTTCCAGCAGAGTTCOCOA 760
 QY 1796 AGCTGGAACGCTAGTGTGCTCTTCTGCTTGAAGTGGCTCCCGGACCGCCAGCGCC 1855
 Db 759 AGCTGGAACGCTAGTGTGCTCTTCTTGAAGTGGCTCCCGGACCGCCAGCGCC 700
 QY 1856 CTCAGTTGCCAGGGATGGGCGCACCACTGTGCACACTGTGGAATACAGACAGTGAATCT 1915
 Db 699 CTCAGTTGCCAGGGATGGGCGCACCACTGTGCACACTGTGGAATACAGACAGTGAATCT 640
 QY 1916 GTCGCTGACGAGTGTGTAATTAAGTCTAGAGCAGCTCTCTAGCAGAGTAAGGT 1975
 Db 639 GTCGCTGACGAGTGTGTAATTAAGTCTAGAGCAGCTCTCTAGCAGAGTAAGGT 580
 QY 1976 CCGCTGACAGTGTGTTGGTGGGCGAGCTTCTGCTCAAAATTCACCAAGCAGAA 2035
 Db 579 CCGCTGACAGTGTGTTGGTGGGCGAGCTTCTGCTCAAAATTCACCAAGCAGAA 520
 QY 2036 TGCCTCTCAGCTCATGTGTGGTCTCTGCTCTCTAGTCTCCCGAGGATGTGGGA 2095
 Db 519 TGCCTCTCAGCTCATGTGTGGTCTCTGCTCTCTAGTCTCCCGAGGATGTGGGA 460
 QY 2096 CCCAGCTGTCTCGGACGTAGAGCAGTACACAGGATGTGGATTTGGGACCTGTGT 2155
 Db 459 CCCAGCTGTCTCGGACGTAGAGCAGTACACAGGATGTGGATTTGGGACCTGTGT 400
 QY 2156 GGTGGCCTTGAAGTGTCTGTGTGTGGAGTCTGCTCCCTTCTTAAAGGAATG 2215

Db 399 GGTGGCCTTGAAGTGTCTGTGTTGTGAGGAGTGAATCCCTTCTTAAAGGAATG 340
 QY 2216 CCCCAGGAGGAGACATTTGGAGGAAGATGGCTGTGAGTGTGACATTTGGCTCTGTACCTG 2275
 Db 339 CCCCAGGAGGAGACATTTGGAGGAAGATGGCTGTGAGTGTGACATTTGGCTCTGTACCTG 280
 QY 2276 CTTCTGAGGCGCCGCTTAAATTAATTCATCAAGATTCCTTTGTAGTTAAAGGTCCTCAGT 2335
 Db 279 CTTCTGAGGCGCCGCTTAAATTAATTCATCAAGATTCCTTTGTAGTTAAAGGTCCTCAGT 220
 QY 2336 TCTGACTGAGGCTCTAGAGAGCTGGGCTTGTATGTTCTTTTGGCCTTTTGTCTACT 2395
 Db 219 TCTGATTTGGAGCTCTAGAGAGCTGGGCTTGTATGTTCTTTTGGCCTTTTGTCTACT 160
 QY 2396 AAATCAAGAAACCATGCTGGAGGGCGGTGAACACAGAACCTCAACAGAGGATGACA 2455
 Db 159 AAATCAAGAAACCATGCTGGAGGGCGGTGAACACAGAACCTCAACAGAGGATGACA 100
 QY 2456 GAGCTGGAGGACACATCTAGCTGCTTGCATTCGCAACCTCACTGGGCTCCCGAGACTCTGTGTG 2515
 Db 99 GAGCTGGAGGACACATCTAGCTGCTTGCATTCGCAACCTCACTGGGCTCCCGAGACTCTGTGTG 40
 QY 2516 TGAGAAATTAACCCCTGCTTGTGTTGAGAAAAA 2554
 Db 39 TGAGAAATTAACCCCTGCTTGTGTTGAAAAA 1

RESULT 7

ABA93718
 ID ABA93718 standard; cDNA; 3071 BP.
 XX ABA93718;
 AC AC
 DT DT
 XX 30-APR-2002 (first entry)
 DE Human transmembrane protein cDNA clone amy2_1i14.
 XX Human; foetal brain; foetal kidney; melanoma; testis; amygdala;
 KW gene therapy; chromosome 22q13.13-13.33; ss.
 XX Homo sapiens.
 OS
 XX WO200198454-A2.
 FN
 XX 27-DEC-2001.
 PD
 XX 25-APR-2001; 2001WO-IB02050.
 PF
 XX 25-APR-2000; 2000US-199380P.
 PR
 XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
 PA
 XX Wiemann S;
 PI
 XX WPI: 2002-055860/07.
 DR P-PSDB; ABB05680.
 DR
 XX Human cDNA sequences and clones derived from human fetal brain, fetal
 PT kidney, melanoma, testis and amygdala cDNA libraries, useful in genetic
 PT screening and therapy -
 XX
 PS Claim 1; Page 128-129; 611pp; English.

XX The present invention describes assemblies and computer readable media
 CC comprising novel human cDNA sequences and clones derived from human
 CC foetal brain, foetal kidney, melanoma, testis and amygdala cDNA
 CC libraries. ABA93702 to ABA93766 represent human cDNA sequences from the
 CC present invention which encode the proteins given in ABB05662 to
 CC ABB05729. The human cDNA sequences and clones can be used in gene
 CC therapy. The clones may be used in a variety of applications, for
 CC example they may be used in profiling assays, for providing large arrays
 CC of human genetic material for implementing large-scale screening
 CC strategies and for treating diseases via gene therapy procedures.

1449 ACCATTAAATCTTTTACAAATATGTGTTGCCACCATAGAAAGGTCTGGCTGATGGAT 1508
 778 ACCTGATGATCTGTGTGACGGGGCCCTCCAGAGATGGCTGGACTGCTGTCTGCTACC 837
 1509 TCCTGATGATGGGATGATGCTCTCAGAGCAGAGATGGATCTGATGCTGTCTGTTATC 1568
 838 ATGCTCTTCCACGGCCATCTTCCGGCCACCTTTCTGTGAGAGAAATGACATGAGCTCA 897
 1569 ATGCAACCTCTCTTCCATTTTCCCTGTGGGTTTCTGTGAAATTAACATGATAGAACTGA 1628
 898 CACGCCAAAGGTTATGAGCAGCAGATTTCACTGGGAGAACTACTTGGAGAGACCA 957
 1629 CTCACCCAGAGGTTACAAAACTTCTTTTAAATGTTTGTACTCTCAGGGAACCG 1688
 958 AGTGAAGAGCGCTCCATCGAGACTCTTTTAACTGATTTGCCAAACCATGGCTTCAAGG 1017
 1689 GCTCATTTGACAGCAGCAGTAAACTATTTTAAAGGATTTTCCAAACCATGGATTCGGT 1748
 1018 TGGGATGAGCTGGAGCCGCTGACCTGATGAGAGCCCGCTCATCTGTGTGCGCCACGG 1077
 1749 TAGGAATGAAATTAAGAGCTGTAGATCTCATGGAGCCACCGTTAATATGTGTAGCCACAG 1808
 1078 TGAACAGAGTGTGATCGGTCTCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACG 1137
 1809 TTACTCGAATTAATCACCATCTCTTGAGATACATTTTGTGTTGGAGAGAGATG 1868
 1138 ACCAGTGGGTGGAGTGGAGTCCCGAGACATCTACCCCGTGGCTGTGTGTGAGCTACCG 1197
 1869 ACCAGTGGGTGAGCTGTGAGTCCCGTACCTCTATCTGTAGGGTGGTGTGAGTTAACTG 1928
 1198 GCTACCGCTCCAGCTCCCTG 1218
 1929 GATATCAACTACGCTCCAG 1949

RESULT 9
 ABX08841
 ID ABX08841 standard, cDNA, 2564 BP.
 XX ABX08841;
 AC
 XX
 XX
 DT 21-JAN-2003 (first entry)
 XX
 DE Angiogenesis-associated human polynucleotide sequence #103.
 XX
 KW Human; angiogenesis-associated transcript; angiogenesis;
 KW angiogenesis-associated disease; cancer; cytostatic; gene therapy;
 KW Gene; ss.
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO200279492-A2.
 XX
 XX 10-OCT-2002.
 XX
 XX 14-FEB-2002; 2002WO-US04915.
 XX
 XX 14-FEB-2001; 2001US-0784356.
 XX 22-FEB-2001; 2001US-0791390.
 XX 19-APR-2001; 2001US-285475P.
 XX 03-AUG-2001; 2001US-31025P.
 XX 13-NOV-2001; 2001US-350666P.
 XX 29-NOV-2001; 2001US-334244P.
 XX
 XX (EOSB-) EOS BIOTECHNOLOGY INC.
 XX
 XX Murray R, Glynn R, Watson SR, Aziz N;
 XX
 XX WPI; 2003-040681/03.
 XX P-PSDB; ABU03557.
 XX
 XX Detecting angiogenesis-associated transcript in a cell for diagnosing
 XX and treating cancer by contacting a sample with a polynucleotide that

PT exhibits changes in expression level as a function of time in tissue
 PT undergoing angiogenesis
 XX
 XX Example 2; Page 274-275; 291pp; English.
 XX
 CC The present invention relates to methods and compositions for
 CC detecting an angiogenesis-associated transcript in a cell in
 CC a patient. The method involves contacting a biological sample from
 CC the patient with a polynucleotide that selectively hybridizes to a
 CC sequence at least 80% identical to any of the angiogenesis-associated
 CC human polynucleotide sequences given in the specification. These
 CC angiogenesis-associated polynucleotide sequences comprise genes that
 CC exhibit changes in expression levels as a function of time in tissue
 CC undergoing angiogenesis. The method and the polynucleotide sequences
 CC of the invention are useful for diagnosing and treating angiogenesis
 CC and angiogenesis-associated diseases e.g. cancer. The polynucleotide
 CC sequences are also useful in the gene therapy of such disorders. The
 CC angiogenesis-associated proteins encoded by the polynucleotide
 CC sequences are useful as a vaccine for therapeutic and prophylactic
 CC immunisation. ABX08739-ABX08853 represent angiogenesis-associated
 CC polynucleotide sequences.
 XX

SQ Sequence 2564 BP; 832 A; 482 C; 560 G; 690 T; 0 other;

Query Match 15.9%; Score 405.2; DB 25; Length 2564;
 Best Local Similarity 60.2%; Pred No. 1.3e-101;
 Matches 721; Conservative 0; Mismatches 453; Indels 24; Gaps 2;

24 TGGGAGGATGTCATGAAAGGGATGAAGGTGGAGGTGCTCAACAGTGATGCTGTGTCGCC 83
 181 TGGGGTGATATCTCAGAAAATGTGAGAGTAGAAGTTCCCAATACAGACTGCGAGCCTACCT 240
 84 AGCCGGGTGCTACTGATCGCCTCTGTCATCCAGACAGCAGGGTATCGGGTGTCTGCTCG 143
 241 ACCAAGTCTTCTGGATTGCTGGAAITGTAATTTAGAGGTTACAATGCCCTTTTAA 300
 144 TATGAAGGCTTTGAAAATGACCCAGCCATGACTTCTGGTGCAACTGGTGGAAACAGTGAT 203
 301 TATGAAGGATTTGAAAATGACTCTGCTCGGACTTCTGGTGCAATATATGTGTTCTGAT 360
 204 GTCCACCCATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 263
 361 ATCCATCCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 420
 264 CATGCCAAGTTCACCGACTGGAAGGGCTACCTCATGAAACGGCTGGTGGTGGTGGTGGTGG 323
 421 CAGCATAAATATACAACTGGAAAGCTTTTCTAGTGAACGACTTACTGTTGCCAAACA 480
 324 CTTCCCGTGGATTTCCATCAAGATGGTGGAGAGCATGAGTACCCCTTTAGGCGAGGC 383
 481 CTGCTCTCTGATTTCTCCAAAAGGTTTCAGAGAGTATGCAAGTATCTCTTCAACCTTGC 540
 384 ATGCGGCTGGAAGTGGTGGACCAAGTCCAGGTGTCCAGCTCTCGCATGGCTGTGGTGGAC 443
 541 ATGAGAGTAGAAGTGGTGGACAGAGGATTTGTTGTCGAACACGAGTAGCAGTGGTGGAA 600
 444 ACAGTAATCGGGGTGCGCTACGGCTCTCTACAGGA ---TGGTGACAGTGACACAGCAG 500
 601 AGTGTAAATTGGAGGAGATTAAGACTAGTGTATGAAGAAAGCCAGATAGAGATAGACATGAC 660
 501 TTCTGGTGCCACATGTGGAGCCCGCTGATCCACCCAGTGGGTGGTGGTGGTGGTGGTGGG 560
 661 TTCTGGTGCCATATGACAGCCCATTAATACATCATATTTGGTGGTGGTGGTGGTGGTGGT 720
 561 CACGGCATCAAGATGTGACAGAGGGGAGTGAAGTGGCCCATCACCCACCTTCCGGAGAG 620
 721 CATCGATTCAAAAGATCTGATATTAACAAGAA-----ACAGGAT 759
 621 ATCTACTGTGATCGCGTCTCTTACCTTCAAGAGGTACGAGCTGTACACAGAGGC 680
 760 GGACATTTTGTATACACCAACCATTTTATTGTTAGGTAAGAGTAGACACAGAGTGGG 819
 681 GGTGTTTGGAGGAGGATGAAGCTGGAGGCCATTGACCCCTGTAATCTGGGCAACATC 740

XX 13-DEC-2001.
XX 30-MAY-2001; 2001WO-US10838.
XX 05-JUN-2000; 2000US-209473P.
XX 05-JUN-2000; 2000US-209531P.
XX 18-SEP-2000; 2000US-233133P.
XX 18-SEP-2000; 2000US-233617P.
XX 20-SEP-2000; 2000US-234009P.
XX 20-SEP-2000; 2000US-234034P.
XX 20-SEP-2000; 2000US-234052P.
XX 22-SEP-2000; 2000US-234509P.
XX 22-SEP-2000; 2000US-234567P.
XX 25-SEP-2000; 2000US-234923P.
XX 25-SEP-2000; 2000US-234924P.
XX 25-SEP-2000; 2000US-235077P.
XX 25-SEP-2000; 2000US-235082P.
XX 25-SEP-2000; 2000US-235134P.
XX 25-SEP-2000; 2000US-235280P.
XX 26-SEP-2000; 2000US-235637P.
XX 26-SEP-2000; 2000US-235638P.
XX 27-SEP-2000; 2000US-235711P.
XX 27-SEP-2000; 2000US-235720P.
XX 27-SEP-2000; 2000US-235840P.
XX 27-SEP-2000; 2000US-235863P.
XX 28-SEP-2000; 2000US-236028P.
XX 28-SEP-2000; 2000US-236032P.
XX 28-SEP-2000; 2000US-236033P.
XX 28-SEP-2000; 2000US-236034P.
XX 28-SEP-2000; 2000US-236109P.
XX 28-SEP-2000; 2000US-236111P.
XX 29-SEP-2000; 2000US-236842P.
XX 29-SEP-2000; 2000US-236891P.
XX 02-OCT-2000; 2000US-237172P.
XX 02-OCT-2000; 2000US-237173P.
XX 02-OCT-2000; 2000US-237278P.
XX 02-OCT-2000; 2000US-237294P.
XX 02-OCT-2000; 2000US-237295P.
XX 02-OCT-2000; 2000US-237316P.
XX 03-OCT-2000; 2000US-237425P.
XX 03-OCT-2000; 2000US-237598P.
XX 03-OCT-2000; 2000US-237604P.
XX 03-OCT-2000; 2000US-237606P.
XX 03-OCT-2000; 2000US-237608P.
XX 01-NOV-2000; 2000US-244867P.
XX 01-NOV-2000; 2000US-245084P.
(AVAL-) AVALON PHARM.
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
XX Soppet DR, Weaver Z;
XX WPI; 2002-188264/24.
XX Screening for anti-neoplastic agent involves exposing cells to a
XX chemical agent to be tested for anti-neoplastic activity, and
XX determining a change in expression of a gene of a signature gene set -
XX
XX Claim 1; SEQ ID 7266; 44pp; English.
XX The present invention describes a method (M1) for screening for an
XX anti-neoplastic agent. The method involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (I)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL61664
XX to ABL70110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (I) has cytostatic
XX activity and can be used in gene therapy. M1 can be used for screening
XX an anti-neoplastic agent, and can be used for producing a product which
XX is the data collected with respect to the anti-neoplastic agent as a
XX result of M1, and the data is sufficient to convey the chemical
XX structure and/or properties of the agent. M1 can be used in the

CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC esophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ Sequence 418 BP; 103 A; 112 C; 124 G; 76 T; 3 other;
Query Match 11.0%; Score 281.4; DB 24; Length 418;
Best Local Similarity 97.3%; Pred. No. 1.2e-67;
Matches 285; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1001 AACCATGGCTTCAAGCTGGGCGCATGAGCTGGAGGCGGTGACCTGATGAGCCCGGCT 1060
DB 1 AACCATGGCTTCAAGCTGGGCGCATGAGCTGGAGGCGGTGACCTGATGAGCCCGGCT 60
QY 1061 CATCTGTGTGGCCACCGGTGAAACGAGTGTGTCATCGGCTCCTCAGCATCCATTTGACGG 1120
DB 61 CATCTGTGTGGCCACCGGTGAAACGAGTGTGTCATCGGCTCCTCAGCATCCATTTGACGG 120
QY 1121 CTGGACAGCGAGTACGACGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1180
DB 121 CTGGACAGCGAGTACGACGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
QY 1181 CTGGTGTGAGCTTACCGGCTACCGGCTACCGGCTCAGCTCAGCTCAGCTCAGCTCAGCT 1240
DB 181 CTGGTGTGAGCTTACCGGCTACCGGCTACCGGCTCAGCTCAGCTCAGCTCAGCTCAGCT 240
QY 1241 GCTCAAGGCCAAAGAGGCCCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1293
DB 241 GCTCAAGGCCAAAGAGGCCCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 293
RESULT 12
AAH87128/c
ID AAH87128 standard; DNA; 289 BP.
XX
AC AAH87128;
XX
DT 27-FEB-2002 (first entry)
XX
DE Human single nucleotide polymorphism containing DNA sequence #1985.
XX
KW Biallelic marker; polymorphism; human; disease; diagnosis; treatment;
KW phenotypic trait; gene therapy; forensic; paternity; mapping; cancer;
KW transgenic; single nucleotide polymorphism; SNP; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Variation replace(126,A)
FT FT /*tag= a
FT FT /standard_name= "single nucleotide polymorphism"
XX
PN WO9953095-A2.
XX
PD 21-OCT-1999.
XX
PF 30-MAR-1999; 99WO-US06893.
XX
PR 09-APR-1998; 98US-0057871.
XX
XX (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Lander ES, Wang D, Hudson T;
XX
DR WPI; 1999-620443/53.
XX
XX Polymorphic human genomic sequences and related allele-specific probes
PT and primers, useful for genetic analysis, e.g. diagnosis and monitoring
PT of disease -
XX
XX Claim 1; Page 243; 330pp; English.
PS

XX CC This invention describes novel human nucleic acid segments (I)
CC containing polymorphic sites. The polynucleotides of (I) are used for,
CC e.g. correlating disease polymorphisms (or disease susceptibility) or
CC other phenotypic traits (e.g. baldness, obesity, fertility, strength,
CC response to drugs etc.); diagnosing and monitoring e.g. cancer,
CC inflammation, heart or central nervous system diseases; detecting
CC susceptibility to microbial infection; treating or preventing such
CC diseases; forensic analysis; gene therapy; paternity testing; mapping
CC genomic loci associated with phenotypic traits (and subsequent cloning
CC of the genes responsible); and the production of transgenic organisms.
CC Antibodies raised against (I) are useful as diagnostic and therapeutic
CC tools and in drug screening. AAH85144 - AAH87644 represent the human
CC DNA sequences containing biallelic polymorphic sites described in the
CC invention.
CC
CC Sequence 289 BP; 72 A; 77 C; 68 G; 72 T; 0 other;
XX QQ

Query Match.	10.6%;	Score 271.8;	DB 20;	Length 289;
Best Local Similarity	98.6%;	Pred. No. 4.6e-65;		
Matches 285;	Conservative 0;	Mismatches 2;	Indels 2;	Gaps 1;
Qy	2180	TTTGTGAGGACTGACTCCC--ATTTCCTAAAGAAATGCCCCGGGAGGACATTGGGAG	2237	
Db	289	TTTGTGAGGACTGACTCCCCATTTCCTAAAGAAATGCCCCGGGAGGACATTGGGAG	230	
Qy	2238	GAAGATGGCCTGAGTGTGCATTGGCTCTGTCTACCTGCTCTCTGAAGCCCGCTAAAAAT	2297	
Db	229	GAAGATGGCCTGAGTGTGCATTGGCTCTGTCTACCTGCTCTCTGAAGCCCGCTAAAAAT	170	
Qy	2298	AAATTCATCCAAGATTCTCTTTGTAGTTAAAGGGTCCAGTTCTGACTGGAGCCTCTAGAGAG	2357	
Db	169	AAATTCATCCAAGATTCTCTTTGTAGTTAAAGGGTCCAGTTCTGACTGGAGCCTCTAGAGAG	110	
Qy	2358	CTGGGCTGTATGTTCTTTTGGCCCTTTTGTCTCTACCTTAATGAAGAAAACCATGCCTTGA	2417	
Db	109	CTGGGCTGTATGTTCTTTTGGCCCTTTTGTCTCTACCTTAATGAAGAAAACCATGCCTTGA	50	
Qy	2418	GGGGCCGTGAACACAGAACCCCTCAAGACAAGGATGACAGAGCTGGAGGA	2466	
Db	49	GGGGCCGTGAACACAGAACCCCTCAAGACAAGGATGACAGAGCTGGAGGA	1	

RESULT 13
ABL27733
ID ABL27733 standard; DNA; 3486 BP.
XX XX ABL27733;
AC AC
DT DT 26-MAR-2002 (first entry)
DE DE
DE DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 34672.
KW KW Drosophila; developmental biology; cell signalling; insecticide;

PT	New isolated nucleic acid detection reagent for detecting 1000 or more	
PT	genes from Drosophila and for elucidating cell signalling and cell-cell	
PT	interactions -	
XX		
PS	Claim 1; SEQ ID NO 34672; 21pp + Sequence Listing; English.	
XX		
CC	The invention relates to an isolated nucleic acid detection reagent	
CC	capable of detecting 1000 or more genes from Drosophila. The invention is	
CC	useful in developmental biology and in elucidating cell signalling and	
CC	cell-cell interactions in higher eukaryotes for the development of	
CC	insecticides, therapeutics and pharmaceutical drugs. The invention	
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA	
CC	sequences (ABL01840-ABL16175) and the encoded proteins	
CC	(ABBS7737-ABBS72072).	
CC	The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published_pct_sequences.	
XX		
SQ	Sequence 3486 BP; 939 A; 901 C; 869 G; 777 T; 0 other;	
	Query Match 10.4%; Score 266.4; DB 23; Length 3486;	
	Best Local Similarity 55.6%; Pred. No. 5e-63;	
	Matches 678; Conservative 0; Mismatches 501; Indels 41; Gaps 7;	
Qy	92 GTACTGGATCGCCTCTGTCATCCAGACAGCAGGGTATCGSGTCTGCTTCGTGTATGAAGG 151	
Db	1198 GTTTTGGTGGCCACCATCTCTGGAATCAAAGGCTATAAGGCCCTTAATGAGCTACGAGGG 1257	
Qy	152 CTTTGAATATGACGCCAGCATGACTCTCTGGTGCACCTGGGAACAGTGGATGTCACCC 211	
Db	1258 TTTTCGATAGGAGTC---GCACGACTTCTGGGTGAACCTCTGCAATGCCGAGGTGCATT 1314	
Qy	212 CATTTGGTGTGTGTCATCAACAGCAGAAATCTAGTGCCTCCACGACCATCCATGCCAA 271	
Db	1315 GTTGGTGTGTGGCCACTCGGGGCAAGCCATTAAATCCGCCCGCACCATCGAGCACAA 1374	
Qy	272 GTTTCACCGACTGAAAGGCTACCTCATGAAAACGCTGTGGGTCCAGACGCTTCCCGT 331	
Db	1375 GTACAAGGACTGAAAGGACTTTCTGTGTGGGAGCTTTATCCGAGCGCCGACCCCTCCCTC 1434	
Qy	332 GGATTTTCCATCAACATGGTGAGAGCATGAAGTACCCTTTTAGCGAGGGATGGGCT 391	
Db	1435 CAATTTTACAAATAATCAACGACAGCTCCAGTCGCGTTCGGCTTGGCTGATCT 1494	
Qy	392 GGAAGTGTGGACAAGTCCCAGTGTCAACGCACTGCATGGCTGTGGTGGACACAGTAAT 451	
Db	1495 CGAGTGGTGGACAAGGATCGAATTTCCGAGGTGCGCTGGCCACCGTCAACAAAATCGT 1554	
Qy	452 CGGGGTGCGCTACGCTCCTCTACAGAGATGGTGACAGTGACGACGACTTCTGGTGCCA 511	
Db	1555 GGGAAAGCGCCTTCTCTGCGCTACTTCGAT-----TCGACGACGGCTTTTGGTGTCA 1608	
Qy	512 CATGTGGAGCCCCCTGATCCACCAGTGGGTGGTCAACGACGTGTGGGCGACGGCAT--- 568	
Db	1609 CGAGGACTGCCCATCATCCATCCAGTTGCTGGGCAACACAGTAGGCGCATTAATCTGGC 1668	
Qy	569 -----CAGATGTGAGAGCGCGAAGTGACATGGCCCATCACCCCACTTCGCGAAGA 621	
Db	1669 TGCACCGCAGGACTATCTGAGCGCATGTAGCTGGTCCGAGGCCATGATTGAGTTCA 1728	
Qy	622 TCTACTGTGATGCGCTTCCTTACTCTTCAAGAAGGTACGAGCAGTCTACACA----- 674	
Db	1729 T-CAGGACGATGCCAATCGAGTTGTTTAAGATGAACTTCACCTTCGAGCAATCTACA 1787	
Qy	675 -----GAGGCGGTTGGTTTGAAGAGGATGAAGCTGAGGCGCATTGACCCCTGA 726	
Db	1788 GTACCGGCAAAACCAATAGCTTTGTGGAGGCGATGAAGCTGGAAGCGGTGGATCCACTCA 1847	
Qy	727 ATCTGGGCAACATCTCGGTGGCAACTGTCTGTAAGGTTCTCTGGATGGATACCTCATGA 786	
Db	1848 ACCTTTTCATCATATGCCCGGCTACAGTAATGGCGGTTCTTAAGTTCGATCATGATGA 1907	
Qy	787 TCTGTGTGGA---CGGGGGGGCCCTCCACAGATGGCTTGGACTGGTTCTGCTACCAATGCCT 843	

1908	Db	TACGCAATTGATTCCTTACCACCGGATGCCTCAGGTCGGATTGGTTCTGTTCACCATGAAA	1967
844	Qy	CTTCCACGCGCATCTTCCCGGCCACCTTCTGTGAGAAGAATGACATTGAGACTCACACCGC	903
1968	Db	AGAGTCGGTGATCTTTCCGGCTGGATTCTGTTCCGTCAACAACATTTGGTTACCCAC	2027
904	Qy	CAAAAGGTTATGAGGCACAGACTTTCAACTGGAGAACTATTTCGGAGAGACCAACTCGA	963
2028	Db	CGAACGGCTACGACTCTCGTACATTCACCTGGAGGGTTACCTCCGGGACACGGGAGCG	2087
964	Qy	AAGCCGCTCCATCGAGACTCTTTAAACATGATTGCCAAACCAATGCTTCAAGTGGGCA	1023
2088	Db	TAGCCGCTGCCAGCATCTATTTTCATCGGATTATTCGGATCATGGATTTCAGTGGGTA	2147
1024	Qy	TGAAGCTGGAGGCGGTGGACCTGATGAGCCCGGCTCACTCTGTGTGGCCACACGGTCAAAAC	1083
2148	Db	TGAGTCTGGAGTGTGCAGATCTCATGGATCCCGACTCGTTTCGGTGGCCACCGTGGCGC	2207
1084	Qy	GAGTGTGCATCGGCTCCTCAGCATCCATTTTCACGCGCTGGGACAGGAGTAGCACCAGT	1143
2208	Db	GAGTGTGTGTGCACACTCAAGGTTCACTTTTCACCGATCGACGGATGAGTAGCACGACGT	2267
1144	Qy	GGGTGGACTCGGAGTCCCCAGACATCTACCCCGTCCGCTGGTGTGAGCTCACCGGCTACC	1203
2268	Db	GGTTGGATTCCGAATCAGCCGATATATATCAAGTCGGATGCTGTACTTGGTCAACCATATA	2327
1204	Qy	AGTCCAG-----CCTCTGTGCGCGCAAGAACCGGCCACACCGCTGAAGGCCAAAGAGGCCA	1260
2328	Db	AGCTAGAGGCCCCACCGAGAGTAGCATCAGCAGGCCCGCCAAACCGGGCACCAAGGCCCA	2387
1261	Qy	CAAAAGAGAAAAAGAAACAG	1280
2388	Db	AAATACAGCGAAAGCGAAAG	2407

RESULT 14
ABLI1985
ID ABL1985 standard; DNA; 4379 BP.
XX
XX ABL1985;
XX
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 11128.
XX
XX DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
XX OS
XX Drosophila melanogaster.
XX
XX PN
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
ET interactions -
XX
XX Claim 1; SEQ ID NO 11128; 2lpp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC

CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC	sequences (ABL01840-ABL16175) and the encoded proteins
CC	(ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 4379 BP; 1162 A; 1151 C; 1118 G; 948 T; 0 other;
	Query Match 10.4%; Score 266.4; DB 23; Length 4379;
	Best Local Similarity 55.6%; Pred. No. 5 se-63;
	Matches 678; Conservative 0; Mismatches 501; Indels 41; Gaps 7;
Qy	92 GTACTGATCGCTCTGTCTCATCCAGACAGCAGGGTATCGGTCGTTCGGTATGAAGG 151
Db	2091 GTTTTGGTGCGCCACCATCTCTGAAATCAAGGCTATAAGGCCCTTAATGAGCTACGAGGG 2150
Qy	152 CTTTGAANAATGACGCAGCCATCACCTTCTGGTGCACCTCGGAACAAGTGATGTCACACC 211
Db	2151 TTTCGATACGGATC---GACAGACTTCTGGGTGAACCTTCGAATGCCAGGTGCATTCT 2207
Qy	212 CATTTGGCTGTGTGCCATCAACAGCAAGATCTTAGTGTCCCCACGGACCATTCATGCCAA 271
Db	2208 GGTGGGTGTGTGGCCACTCGGGGCAAGCCATAAATTCGCCGCCGCAACCATCGAGCAA 2267
Qy	272 GTTCACCGACTGGAAGGGCTACTCATGAACCGCTGTGGGTCTCCAGGACGCTTCCCCT 331
Db	2268 GTACAGGACTGGAAGACTTTCTGTGTGGACGTTTTATCCGGAGCCCGCACCTTCCCTC 2327
Qy	332 GGATTTCCACATCAAGATGTGTGAGAGCATGAAGTACCCCTTTAGCGAGGGCATGCCGT 391
Db	2328 CAACCTTTTACAACAATAACAGACGCTCCAGTCGCGCTTCGCGCTTGGCCTGAATCT 2387
Qy	392 GGAAGTGTGGAAGAAGTCCAGGTGTCAACGACTCGCATGGCTGTGTGGAGACACAGTAAT 451
Db	2388 CGAGTCGTGGACAAAGATCGCATTTCCGAGGTGCGCTCGCCACCCTCACAAAATCGT 2447
Qy	452 CGGGGTGTGGCTACGGCTCTCTACGAGGATGTGACAGTGAACGACACTTCTGGTGCCA 511
Db	2448 GGGAAAGCGCTCTTCTCGCGTACTTCGAT-----TCGACGACGGCTTTTGGTGCA 2501
Qy	512 CATGTGAGGCCCTGTATCAACCAGTGGGTGGTTCACGAGCTGTGGGCCACCGCAT---- 568
Db	2502 CGAGGACTCGCCCATCATCTCATCGATTGCTGGGCAACCAACAGTAGGCCATTAATCTGGC 2561
Qy	569 -----CAAGATGTCAGAGAGGCGAAGTGACATGGCCCATCACCCACACTTCGCGAAGA 621
Db	2562 TGCACCGCAGGACTATCTCGAGCGCATGTTAGCTGGTTCGCGAAGCCATGATTGAGGTCA 2621
Qy	622 TCCTACTGTGATCGCGTTCCTTACCTCTTCAAGAGGTTACGACGAGTCTACACA----- 674
Db	2622 T-GAGGACGATGCCACAATCGAGTTGTTTAAGATGAACCTTCACCTTCGACGAATCTACA 2680
Qy	675 -----GAAGGCGGTGTGTTTCAGGAAAGGATGAAGCTGGAGGCCCAITGACCCCTCGA 726
Db	2681 GTGACGCAAAACCAATAGCTTTGTGAGGGCATGAAGCTGGAAGCGGTGATCCACTCA 2740
Qy	727 ATCTGGGCAACATCTCGGTGGCAACTGTCTGTAAAGTTTCTCTGGATGGATACCTGATGA 786
Db	2741 ACCTTTTCATCCATATGCCCGGCTACAGTAATGCGGTTCTTAAAGTTCGGATACATGATGA 2800
Qy	787 TCTGTGTGGA---CGGGGGGCGCTCCACAGATGGCTTGGACTGGTTCCTGCTACCATGCC 843
Db	2801 TAGCATTGATCTTACCAACCGGATGCCTCAGGGTCGGATGGTCTGTGTACCATGAA 2860
Qy	844 CTTCACAGCCOATCTTCGGGCCACCTTCTGTTCAGAGAAGATGAATGAGCTCAACCGC 903
Db	2861 AGAGTCGCTGTATCTTTTCGGCTGGATCTCTGTTCCGTCACAAACAAATTCGGTTACCCAC 2920

QY 904 CAAAGGTTATAGGACAGACATTTCAACTGGGAGAACTACTTTGGAGAAACCAAGTCGA 963
 Db 2921 CGAACGCTACGACTCTCGTACATTCACCTGGGAGGGTTACCTCCGACACGGAGCGG 2980
 QY 964 AAGCCGCTCATCGAGACTCTTTAAACATGATGTCGCAACCATGCTTCAAGTGGGCA 1023
 Db 2981 TAGCCGCTGGCCAGCATCTTTATTCATCGGATTAATCCCGATCATGATTTGAGGTGGTA 3040
 QY 1024 TGAAGCTGGAGGCGGTGACCTGATGAGCCCGGCTCATCTGTGTGGCCACCGTGAAC 1083
 Db 3041 TGATCTGGAGTGGCAGATCTCATGATCCCGACTCGTTTGGTGGCCACCGTGGCG 3100
 QY 1084 GAGTGGTCATCGGCTCTCAGCATCCATTTGACGCTGGGACAGGATGACACAGT 1143
 Db 3101 GAGTGGTTGGTTCGACTACTCAAGGTTTCACTTACCGATGGAGCGATGATGACGACGT 3160
 QY 1144 GGTGGACTGCGACTCCCGACATCTACCCGCTGGTGTGAGCTCACCGGTACC 1203
 Db 3161 GGTGGATTGGCAATCAGCCGATATATTCAGTCGATGGTGTGTACTGTCAACATA 3220
 QY 1204 AGCTCCAG---CCTCCTGTGGCCGACAGACCGGCCACACCGCTGAAGGCCAAAGAGGCCA 1260
 Db 3221 AGCTAGAGGGCCCAACCGAGAGTAGCATCATCAGCAGGCCCGGAAACCGCACCAAGGCCA 3280
 QY 1261 CAAGAGAGAAAGAAACAG 1280
 Db 3281 AATACAGCGAAAGCGAAAG 3300

RESULT 15

AAT23623
 ID AAT23623 standard; cDNA to mRNA; 290 BP.

AC AAT23623;
 XX

DT 05-SEP-1996 (first entry)
 XX

DE Human gene signature HUMG905481.
 XX

XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.
 XX

OS Homo sapiens.
 XX
 FN W09514772-A1.
 XX

PD 01-JUN-1995.
 XX

PF 11-NOV-1994; 94WO-JP01916.
 XX

PR 12-NOV-1993; 93JP-0355504.
 XX

XX (MATS/) MATSUBARA K.
 PA (OKUB/) OKUBO K.
 XX

PI Matsubara K, Okubo K;
 XX

DR WPI; 1995-206931/27.
 XX

XX Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues

XX Claim 1; Page 1414; 2245pp; Japanese.
 PS

XX A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in AAT19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.
 XX

SQ Sequence 290 BP; 46 A; 90 C; 75 G; 64 T; 15 other;

Query Match 9.9%; Score 252.8; DB 16; Length 290;
 Best Local Similarity 92.7%; Pred. No. 8.6e-60; Indels 1; Gaps 1;
 Matches 268; Conservative 0; Mismatches 20;

QY 1619 GATCCTCTCTGTGTAATAATCTGCCCGGTCTGTGAAGGCTGGACGCTGGAGACCTGTG 1678
 Db 1 GATCCTCTCTGTGTAATAATCTGCCCGGTCTGTGAAGGCTGGACGCTGGAGACCTGTG 60
 QY 1679 GGGTCTCTGGACCCGCGCTGTGTTCTGCCCTCCCTGTGGAAGGCTCTATATGACGG 1738
 Db 61 GGGTCTCTGGNACCCGCGCTGTGTTCTGCCCTCCCTGTGGAAGGCTCTATATGACGG 120
 QY 1739 GCCGCTTGAGGCCCCAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCCAAGC 1798
 Db 121 GCTGCTGAGGCCCCAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCCAAGC 180
 QY 1799 TGGAAAGCTAGCTGCTGCTCTTCTTAAGATGGGCTCCCGCCGACCGCCCTC 1858
 Db 181 TGGAAAGCTAGCTGCTGCTCTTCTTAAAGTGGCTCCCGCCGACCGCCCTC 240
 QY 1859 AGTTGCCAGGGATGGGCCA-CCACTGTCACTGTGGAATACAGACA 1906
 Db 241 AGTTGCCAGGGCTGGGCCAACCCTGTACACTGNGGNTGCANGANA 289

Search completed: February 4, 2004, 13:46:59
 Job time : 691 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 4, 2004, 13:26:59 ; Search time 9185 Seconds
(without alignments)
11379.869 Million cell updates/sec

Title: US-10-031-915-90
Perfect score: 2555
Sequence: 1 agacgcccactctatgac.....ttgcttgagaaaaaaa 2555

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pi.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2555	100.0	2555	6	AX078286	Sequence
2	2553.4	99.9	3194	9	HSA305226	AX078286 Homo sapi
3	2553.4	99.9	3222	9	BC017191	AX078286 Homo sapi
4	2538.8	98.4	3188	6	AX714233	Sequence
5	2538.8	98.4	3188	9	AX056443	Homo sapi
6	2438.4	95.4	3299	9	HSA305227	AX056443 Homo sapi
7	2046.8	80.1	5875	9	AK074091	AX056443 Homo sapi
8	1434.6	58.1	3071	9	HSB01538	AX074091 Homo sapi
9	1227.8	46.1	2432	9	AK097052	AX074091 Homo sapi
10	1203.6	47.1	3395	10	BC030864	BC030864 Mus muscu
11	1132.4	44.3	8948	9	HS75623	AL035681 Human DNA
12	1128.2	44.2	3399	10	AY237001	AY237001 Mus muscu
13	867.4	33.9	3725	10	AY237002	AY237002 Mus muscu
14	454.8	17.8	458	6	BD115452	BD115452 EST and e
15	413.8	16.2	3590	6	AX046388	AX046388 Sequence
16	410.6	16.1	2878	10	BC020018	BC020018 Mus muscu
17	405.2	15.9	2564	9	AK000062	AK000062 Homo sapi
18	324.4	12.7	400	11	GI4475	GI4475 human STS S
19	281.4	11.0	418	6	AX336757	AX336757 Sequence
20	270.8	10.6	289	11	G43605	G43605 WIAF-2484-S
21	266.4	10.4	4809	3	BT006011	BT006011 Drosophil
22	236.4	9.3	250	11	GI4931	GI4931 human STS S
23	234.8	9.2	164681	2	AC101996	AC101996 Mus muscu
24	210.6	8.2	193660	2	AC102262	AC102262 Mus muscu
25	204.6	8.0	251144	2	AC109565	AC109565 Rattus no
26	204.6	8.0	258549	2	AC133113	AC133113 Rattus no
27	158.8	6.2	174036	3	AC009200	AC009200 Drosophil
28	158.8	6.2	174036	3	AC010214	AC010214 Drosophil
29	158.8	6.2	261775	3	AE003639	AE003639 Drosophil
30	144.6	5.7	199301	2	EX088589	EX088589 Danio rer
31	135.8	5.3	165	6	AX209757	AX209757 Sequence
32	133.6	5.2	2444	9	BC039820	BC039820 Homo sapi
33	133.6	5.2	3394	9	HS089358	HS089358 Homo sapi
34	133.4	5.2	4323	9	AB014581	AB014581 Homo sapi
35	129.2	5.1	62093	2	AC099875	AC099875 Mus muscu
36	127.8	5.0	215435	10	AL662838	AL662838 Mouse DNA
37	124.2	4.9	3315	9	HSB00860	AL110279 Homo sapi
38	122.4	4.8	62093	2	AC099875	AC099875 Mus muscu
39	121.4	4.8	128328	9	AC005839	AC005839 Homo sapi
40	121	4.7	612	9	BC012882	BC012882 Homo sapi
41	121	4.7	1570	9	BC048421	BC048421 Homo sapi
42	121	4.7	2285	9	BC040188	BC040188 Homo sapi
43	117.8	4.6	2636	6	AX156560	AX156560 Sequence
44	102.6	4.0	2231	10	BC036972	BC036972 Mus muscu
45	102.6	4.0	2781	10	AB032165	AB032165 Mus muscu

ALIGNMENTS

RESULT 1
AX078286
LOCUS AX078286
DEFINITION Sequence 90 from Patent WO0107471.
ACCESSION AX078286
VERSION AX078286.1 GI:13157977
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Hillman,J.L., Lal,P., Tang,Y.T., Yue,H., Au-Young,J., Bandman,O.,
Azimzai,Y., Yang,J., Lu,D.A., Baughn,M.R., Patterson,C. and Shah,P.
TITLE Cell cycle and proliferation proteins

AX078286 linear PAT 22-FEB-2001

JOURNAL	Patent: WO 0107471-A 90 01-FEB-2001;
FEATURES	Incyte Genomics, Inc. (US)
source	Location/Qualifiers
	1..2555
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/note="Incyte ID No: 1868749CB1"
BASE COUNT	577 a 709 c 722 g 547 t
ORIGIN	
	Query Match 100.0%; Score 2555; DB 6; Length 2555;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 2555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 AGCAGTCCCACTCTATGACCAAGTGGAGGATGTGATGAAAGGATGAAGGTGGAGGTGC 60
DB	1 AGCAGTCCCACTCTATGACCAAGTGGAGGATGTGATGAAAGGATGAAGGTGGAGGTGC 60
QY	61 TCAACAGTACGCTGCTCCCCAGCGGGTGACTCGATCGCCTCTGTCAATCCACGACAG 120
DB	61 TCAACAGTACGCTGCTCCCCAGCGGGTGACTCGATCGCCTCTGTCAATCCACGACAG 120
QY	121 CAGGGTATCGGTGCTGCTTCGGTATGAAGGCTTTGAAATGACGCCACGCCATGACTTCT 180
DB	121 CAGGGTATCGGTGCTGCTTCGGTATGAAGGCTTTGAAATGACGCCACGCCATGACTTCT 180
QY	181 GGTGCMACTGGGAACAGTGAATGCCACCCATTGGCTGGTGTGCATCAACAGCAGA 240
DB	181 GGTCAACCTGGGAACAGTGAATGCCACCCATTGGCTGGTGTGCATCAACAGCAGA 240
QY	241 TCTAGTGGCCCCACGGACCATCCATGCAAGTTCCAGCACTGGAAAGGGTCACTCATGA 300
DB	241 TCTAGTGGCCCCACGGACCATCCATGCCAGTTCAACGACTGGAAAGGGTCACTCATGA 300
QY	301 AACCGCTGTGGGCTCCAGGACGCTTCOCGPGGATTTCCATCAAGATGGTGGAGAGA 360
DB	301 AACCGCTGTGGGCTCCAGGACGCTTCOCGPGGATTTCCATCAAGATGGTGGAGAGA 360
QY	361 TGAAGTACCCCTTTAGGCAGGGCATGCGGCTGGGAAGTGGTGGACAAGTCCCAAGTGTCTAC 420
DB	361 TGAAGTACCCCTTTAGGCAGGGCATGCGGCTGGGAAGTGGTGGACAAGTCCCAAGTGTCTAC 420
QY	421 GCACTCGCATGGCTGTGGTGGACACAGTAATTCGGGGGTCGCCCTACGGCTCTCTACGAG 480
DB	421 GCACTCGCATGGCTGTGGTGGACACAGTAATTCGGGGGTCGCCCTACGGCTCTCTACGAG 480
QY	481 ATGGTGAAGTGAAGACAGCACTCTGGTGCACATGTGGAGCCCTGTATCCACCCAGTGG 540
DB	481 ATGGTGAAGTGAAGACAGCACTCTGGTGCACATGTGGAGCCCTGTATCCACCCAGTGG 540
QY	541 GTTGGTCAAGAGCTGTGGGCCACGGCATCAAGATGTCAAGAGCGCGAAGTGCATGGCCC 600
DB	541 GTTGGTCAAGAGCTGTGGGCCACGGCATCAAGATGTCAAGAGCGCGAAGTGCATGGCCC 600
QY	601 ATCACCCCACCTTCGGGAAGATCTTACTGTGAAGCGGTTCCTTACTCTTCAAGAAGGTAC 660
DB	601 ATCACCCCACCTTCGGGAAGATCTTACTGTGAAGCGGTTCCTTACTCTTCAAGAAGGTAC 660
QY	661 GAGCACTCTACACAGAACGGGTGGTTTGGAGGAGGATCAAGCTGGAGGCCATTGACC 720
DB	661 GAGCACTCTACACAGAACGGGTGGTTTGGAGGAGGATCAAGCTGGAGGCCATTGACC 720
QY	721 CCCTGAACTCGGCAACATCTCGTGGCAACTGTCTGTAAAGTTCTCTGGATGGATACC 780
DB	721 CCCTGAACTCGGCAACATCTCGTGGCAACTGTCTGTAAAGTTCTCTGGATGGATACC 780
QY	781 TGATGATCTGTGTGACGCGGGGGCCCTCCACAGATGGCTTGGACTGGTCTCTGTACCATG 840
DB	781 TGATGATCTGTGTGACGCGGGGGCCCTCCACAGATGGCTTGGACTGGTCTCTGTACCATG 840
QY	841 CCTCTTCCACGCGCATCTTCCCGGGCCACTTCTGTCAAGAAGATGCAATTGACTCACCAC 900

841	DB	CCTCTTCCACGCGCATCTCTCCCGGCCACCTTCTGTGTCAGAAAGATGACATTTGAGCTCACAC	900
901	QY	CGCAAAAGGTTATAGGGCAGACACTTTCAACTGGGAGAACTACTTTGGAGAGACCAAGT	960
901	DB	CGCAAAAGGTTATAGGGCAGACACTTTCAACTGGGAGAACTACTTTGGAGAGACCAAGT	960
961	QY	CGAAAGCGCTTCATCGAGACTCTTTTAAATGATGATTTGCCCAACCATTGGCTTCAAGGTGG	1020
961	DB	CGAAAGCGCTTCATCGAGACTCTTTTAAATGATGATTTGCCCAACCATTGGCTTCAAGGTGG	1020
1021	QY	GCATGAAGCTGGAGGCGGTGGACCTGATGAGGCCCGCGGCTCATCTGTGTGGCCACCGGTGA	1080
1021	DB	GCATGAAGCTGGAGGCGGTGGACCTGATGAGGCCCGCGGCTCATCTGTGTGGCCACCGGTGA	1080
1081	QY	AACGAGTGTGTCATCGGCTCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACGACC	1140
1081	DB	AACGAGTGTGTCATCGGCTCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACGACC	1140
1141	QY	AGTGGGTGGACTCGGAGTCCCGAGACATCTACCCCGTGGCTGGTGTGAGCTCAACCGGCT	1200
1141	DB	AGTGGGTGGACTCGGAGTCCCGAGACATCTACCCCGTGGCTGGTGTGAGCTCAACCGGCT	1200
1201	QY	ACCAGCTCCAGCCTCTCTGTGGCCGCGAGAACCGGCCACACCGCTGAAGGCCAAAGAGGCCA	1260
1201	DB	ACCAGCTCCAGCCTCTCTGTGGCCGCGAGAACCGGCCACACCGCTGAAGGCCAAAGAGGCCA	1260
1261	QY	CAAGAAGAAAAGAAACAGTTTGGAGAGAAAGGAAAGAAATCCGCCCACTAAAGACGC	1320
1261	DB	CAAGAAGAAAAGAAACAGTTTGGAGAGAAAGGAAAGAAATCCGCCCACTAAAGACGC	1320
1321	QY	GACCCCTCAGACAGGGGTCCAGAAGCCCTGTGTGGAGGACGACCCCTCAGGGTGCCAGGA	1380
1321	DB	GACCCCTCAGACAGGGGTCCAGAAGCCCTGTGTGGAGGACGACCCCTCAGGGTGCCAGGA	1380
1381	QY	AGATCTGTCGGAGCCTGTTCCTGGCGAGATCAATGCTGTGCGTGTGAAAGGAGAGATC	1440
1381	DB	AGATCTGTCGGAGCCTGTTCCTGGCGAGATCAATGCTGTGCGTGTGAAAGGAGAGATC	1440
1441	QY	TAGACGTGGCTCTCGCCGACAAAGCTTCAAGTCCAGAGCTGCCTGTCTCCGTCCAGAAACA	1500
1441	DB	TAGACGTGGCTCTCGCCGACAAAGCTTCAAGTCCAGAGCTGCCTGTCTCCGTCCAGAAACA	1500
1501	QY	TCAAGCAGGAAAACAGACGATGAGCCTTCTGCTCCAGCCTGGCTTCTAGCTGGAGACC	1560
1501	DB	TCAAGCAGGAAAACAGACGATGAGCCTTCTGCTCCAGCCTGGCTTCTAGCTGGAGACC	1560
1561	QY	AGCCACGCTTCTCTACACACCAACCAATGSCCTCCACCTGACTTTCGCTTGGAGACTGA	1620
1561	DB	AGCCACGCTTCTCTACACACCAACCAATGSCCTCCACCTGACTTTCGCTTGGAGACTGA	1620
1621	QY	TCCTCTCTGTGTAAATTTCTCCCGGTGCTGTGAAGGCTGGAACGCTGAGGACCTGTCTGGG	1680
1621	DB	TCCTCTCTGTGTAAATTTCTCCCGGTGCTGTGAAGGCTGGAACGCTGAGGACCTGTCTGGG	1680
1681	QY	GTCTCTGGGACCGCCTGTGCTTCTGCGCTCCCTGTGGAAAGTCTATATCACGGGC	1740
1681	DB	GTCTCTGGGACCGCCTGTGCTTCTGCGCTCCCTGTGGAAAGTCTATATCACGGGC	1740
1741	QY	CGCTGAGGCCCCAGAACTCGTCTGTGAACCAACCTTTTCCAGCCAGGATTTCCCAAGAGCTG	1800
1741	DB	CGCTGAGGCCCCAGAACTCGTCTGTGAACCAACCTTTTCCAGCCAGGATTTCCCAAGAGCTG	1800
1801	QY	GAACGCTAGCTGCTCTTCTTTAAGATGGCCTCCCGCGACCCGCGACGGGCCCTCAG	1860
1801	DB	GAACGCTAGCTGCTCTTCTTTAAGATGGCCTCCCGCGACCCGCGACGGGCCCTCAG	1860
1861	QY	TTGCCACGGATGGGCCCAACACTGTACACTGTGGAATAAAGACAGTGAACCTGTGCTG	1920
1861	DB	TTGCCACGGATGGGCCCAACACTGTACACTGTGGAATAAAGACAGTGAACCTGTGCTG	1920
1921	QY	CCTGAAACGAGTCTATGTAAATTAAGTTCTAGACGAGCTCTCTGTAGCAGGATAGGTCCCT	1980
1921	DB	CCTGAAACGAGTCTATGTAAATTAAGTTCTAGACGAGCTCTCTGTAGCAGGATAGGTCCCT	1980

QY	1981	GACAGTCAGTTGTGTGTGGGGGAGGCTCTGCTCAAAAATTCAACAAGCAGAAATGCCT	2040
Db	1981	GACAGTCAGTTGTGTGTGGGGGAGGCTCTGCTCAAAAATTCAACAAGCAGAAATGCCT	2040
QY	2041	CTCAGCCTCATGTGTTGGTCCTCTGCTCTCTCTAGCTCCCAAGGATGTTGGGGACCCAG	2100
Db	2041	CTCAGCCTCATGTGTTGGTCCTCTGCTCTCTCTAGCTCCCAAGGATGTTGGGGACCCAG	2100
QY	2101	CTTGTCTGGCAGCTAAGAAGCAGTGCACAGGATGTGGATTTTGGGACCTGTGTGGTGG	2160
Db	2101	CTTGTCTGGCAGCTAAGAAGCAGTGCACAGGATGTGGATTTTGGGACCTGTGTGGTGG	2160
QY	2161	CCTTGAGCTGCTTTCTGTGTTGTGAGGACTGACTCCCATTTCTTAAAGAAATGCCCC	2220
Db	2161	CCTTGAGCTGCTTTCTGTGTTGTGAGGACTGACTCCCATTTCTTAAAGAAATGCCCC	2220
QY	2221	GGGAGGACATTTGGGAGAAAGATGGGCTGAGTGTGCATTTTGGCTCTGTCTACCTGTCTCT	2280
Db	2221	GGGAGGACATTTGGGAGAAAGATGGGCTGAGTGTGCATTTTGGCTCTGTCTACCTGTCTCT	2280
QY	2281	GAAGCCCCGTAAAAATAATTCATCCAAGATTCCTTGTAGTTAAAGGTCACAGTTCTGA	2340
Db	2281	GAAGCCCCGTAAAAATAATTCATCCAAGATTCCTTGTAGTTAAAGGTCACAGTTCTGA	2340
QY	2341	CTGGAGCCTCTAGAGAGCTGGGCTGTATGTTCTTTTGGCTTTTGTCTACCTAAATG	2400
Db	2341	CTGGAGCCTCTAGAGAGCTGGGCTGTATGTTCTTTTGGCTTTTGTCTACCTAAATG	2400
QY	2401	AAGAAACCATGCTGGAGGGGCGTGNACACAGAACCTCAAGACAAGGATGACAGACT	2460
Db	2401	AAGAAACCATGCTGGAGGGGCGTGAACACAGAACCTCAAGACAAGGATGACAGACT	2460
QY	2461	GGAGGACACATCTAGCTGCCATTGCAACCTCACTGGGCTCCCCAGACTCTGTGTGTGAGA	2520
Db	2461	GGAGGACACATCTAGCTGCCATTGCAACCTCACTGGGCTCCCCAGACTCTGTGTGTGAGA	2520
QY	2521	AATTAAACCCCTGCTTGTGTGAGAAAAA	2555
Db	2521	AATTAAACCCCTGCTTGTGTGAGAAAAA	2555

```

RESULT 2
HSA305226
LOCUS
DEFINITION
HSA305226      3194 bp      mRNA      linear      PRI 18-JAN-2002
Homo sapiens mRNA for H-1(3)mbt-like protein, alternative variant
a.
ACCESSION
AJ305226
VERSION
AJ305226.1      GI:113940238
KEYWORDS
alternative splicing; H-1(3)mbt-like gene; H-1(3)mbt-like protein.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
Wismar, J.
Molecular characterization of h-1(3)mbt-like: a new member of the
human mbt family
PEBS Lett. 507 (1), 119-121 (2001)
21538645
MEDLINE
11682070
PUBMED
REFERENCE
2
Wismar, J.
Direct Submission
Submitted (25-JAN-2001) Wismar J., Johannes Gutenberg Universitaet,
Institut fuer Genetik, Becherweg 32, 55099 Mainz, GERMANY
Location/Qualifiers
1..3194
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="22"
/map="22q13.31-33"
FEATURES
source

```


Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;
 Web site: <http://www.nisac.nih.gov/>
 Contact: nisc_mgc@hgrl.nih.gov
 Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J., Tionsgon, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 4 Row: b Column: 5
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13899349.

FEATURES

source	Location/Qualifiers
1..3222	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/db_xref="taxon:9606"
	/clone="MGC:2476 IMAGE:313844"
	/tissue_type="Placenta, Choriocarcinoma"
	/clone_lib="NIH_MGC_21"
	/lab_host="DH10B-R"
	/note="Vector: pOTB7"
53..2170	/codon_start=1
	/product="hypothetical protein DKFZp7611141"
	/protein_id="AAH17191.1"
	/db_xref="GI:16877935"
	/translation="MEXPRISIEETPSPEEEDDLLEFGYDVSFRSYNSVSGES SYLSESEANEDEARELPTSLHLLSPGTPSLDGSSEPAVEMCIGIVTRAF SFYKTRFVSVCSYNSKASILARLOGKPTKAKVHLKAAASAKIGALFHSQF TQGLADPTQDQALVLPDQKELKDSYKAAVSCFKHVPYLDQMEDVWGMKVEV LNSDAVPSRVYASVITQAGYRVLRYEGFENDASHDFWNLGTVDVHPICWALN SKILVPRTHIAKFTDWKYLMLKELVSRILPVDHFKWVESMKYIPFRGMRLVVDK SQVSRMAVVDVIGGLRLLEDGSDDDFWCHWSPLIHPVSRWSRVGHGKWS RRSDMAHPTPRKTCYDAVYLFKXRAVYEGWFEEMKLEAIDPLNLGNICVATV CKVLIDGYLIMICVDPGSPDGLDFWFCYHASSHAIFFATPCQKNDIELTPPKYEATF NWNVLEKTSKAAPSLENDPCPHGFKVGMKLEAVDLMEPRLICVATVRYVHRL SIHEDGWSEYDQVDCESPIYVPGCELTGYLOPPVAEPATPLKAEATKKKK SIFGKKRIPKTRPQSGSKLLEDDPQGAARKISSEVPVGEIIVRVKEHLDVA SPDKASPELPPSVENIKQETDD"
BASE COUNT	756 a 870 c 913 g 683 t
ORIGIN	
Query Match	99.9%; Score 2553.4; DB 9; Length 3222;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 2554; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY	1 AGCAGTCCCACTCTATGACAGTGGGAGGATGTGATGAAAGGGATGAAGTGGAGTGC 60
DB	648 AGCAGTCCCACTCTATGACAGTGGGAGGATGTGATGAAAGGGATGAAGTGGAGTGC 707
QY	61 TCACAGTGTGCTGTGCTCCAGCGGGTGTACTGATCGCCTCTGTATCCAGACAG 120
DB	708 TCACAGTGTGCTGTGCTCCAGCGGGTGTACTGATCGCCTCTGTATCCAGACAG 767
QY	121 CAGGATATCGGGTGTGCTTCCGATGAAAGCTTTGAAATGACGCCAGGCATGACTTCT 180
DB	768 CAGGATATCGGGTGTGCTTCCGATGAAAGCTTTGAAATGACGCCAGGCATGACTTCT 827
QY	181 GGTGCAACCTGGGAAACAGTGGATGTCACCCCATTTGGTGTGTGCTCCATCAACAGCAAGA 240
DB	828 GGTGCAACCTGGGAAACAGTGGATGTCACCCCATTTGGTGTGTGCTCCATCAACAGCAAGA 887
QY	241 TCTAGTGTGCCCCCAGGACCATCCATGCCAGTTCACGACGTGGAGGGCTACCTCATGA 300
DB	888 TCTAGTGTGCCCCCAGGACCATCCATGCCAGTTCACGACGTGGAGGGCTACCTCATGA 947
QY	301 AACGGTGTGGGCTCCAGGACGCTTCCCGTGGATTTCCACATCAAGATGTGGAGAGCA 360
DB	948 AACGGTGTGGGCTCCAGGACGCTTCCCGTGGATTTCCACATCAAGATGTGGAGAGCA 1007
QY	361 TGAAGTACCCCTTTAGGACAGGCATCGGCTGGAGTGTGACAAAGTCCAGAGTGTAC 420
DB	1008 TGAAGTACCCCTTTAGGACAGGCATCGGCTGGAGTGTGACAAAGTCCAGAGTGTAC 1067
QY	421 GCACTCGCATGCTGTGGTGGACACAGTAATCGGGGGTCCGCTACGGGTCTCTCTACGAGG 480
DB	1068 GCACTCGCATGCTGTGGTGGACACAGTAATCGGGGGTCCGCTACGGGTCTCTCTACGAGG 1127
QY	481 ATGGTGACAGTGAACGACACTTCTGTGTGATGCGCTTCTTACCTCTTCAAGAGGATAC 540
DB	1128 ATGGTGACAGTGAACGACACTTCTGTGTGATGCGCTTCTTACCTCTTCAAGAGGATAC 1187
QY	541 GTTGGTACAGCTGTGGGCCACGACATCAAGATGTCAGAGAGGCGAAGTGAATGACATGGCCC 600
DB	1188 GTTGGTACAGCTGTGGGCCACGACATCAAGATGTCAGAGAGGCGAAGTGAATGACATGGCCC 1247
QY	601 ATACCCCACTTCCGGAGAGTCTACTGTGATGCGCTTCTTACCTCTTCAAGAGGATAC 660
DB	1248 ATACCCCACTTCCGGAGAGTCTACTGTGATGCGCTTCTTACCTCTTCAAGAGGATAC 1307
QY	661 GAGCAGTCTACAGAGAGCGCTTGGTTTGGAGAGGATGAAGCTGGAGGCCATGATGACC 720
DB	1308 GAGCAGTCTACAGAGAGCGCTTGGTTTGGAGAGGATGAAGCTGGAGGCCATGATGACC 1367
QY	721 CCTGAACTCGGGCAACATCTCGTGGCAACTGTCTGTAAGGTTCTCTGATGATGATACC 780
DB	1368 CCTGAACTCGGGCAACATCTCGTGGCAACTGTCTGTAAGGTTCTCTGATGATGATACC 1427
QY	781 TGATGATCTGTGGAGCGGGGGCCCTCCACAGATGGCTTGGAGTGTGCTTCTGCTACCATG 840
DB	1428 TGATGATCTGTGGAGCGGGGGCCCTCCACAGATGGCTTGGAGTGTGCTTCTGCTACCATG 1487
QY	841 CCTCTTCCACGCCACTTTCGCCGCCACCTTCTGTGACAGAAATGACATTTGAGCTCACAC 900
DB	1488 CCTCTTCCACGCCACTTTCGCCGCCACCTTCTGTGACAGAAATGACATTTGAGCTCACAC 1547
QY	901 CGCCAAAAGGTTATGAGGCACAGACTTTCAACTGGGAGAACTTCTGGAGAGACCAAGT 960
DB	1548 CGCCAAAAGGTTATGAGGCACAGACTTTCAACTGGGAGAACTTCTGGAGAGACCAAGT 1607
QY	961 CGAAGCGGCTCCATCGAGACTCTTTTAAATGATGATTTGCCAAACCATGGCTTCAAGGTGG 1020
DB	1608 CGAAGCGGCTCCATCGAGACTCTTTTAAATGATGATTTGCCAAACCATGGCTTCAAGGTGG 1667
QY	1021 GCATGAAGCTGGAGCGCGTGGACCTGTGAGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1080
DB	1668 GCATGAAGCTGGAGCGCGTGGACCTGTGAGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1727
QY	1081 AACGAGTGTGATCGGCTCTCAGATTCACATTTTGAAGCTGGGACAGCAGATGAGGAC 1140
DB	1728 AACGAGTGTGATCGGCTCTCAGATTCACATTTTGAAGCTGGGACAGCAGATGAGGAC 1787
QY	1141 AGTGGTGTGACTCGGAGTCCCGACAGATCTACCCGCTGGCTGGTGTGAGCTCACCGGCT 1200
DB	1788 AGTGGTGTGACTCGGAGTCCCGACAGATCTACCCGCTGGCTGGTGTGAGCTCACCGGCT 1847
QY	1201 ACCAGTCCAGCTCTCTGTGGCCGAGAACCGGCGCACACCGCTGAAGGCGCAAGAGGCGCA 1260
DB	1848 ACCAGTCCAGCTCTCTGTGGCCGAGAACCGGCGCACACCGCTGAAGGCGCAAGAGGCGCA 1907

REMARK
 COMMENT

[illegible]

Db	2988	TTGAGGCTCTAGAGAGCTGGGCTTGATGTTCTTTTGGCTTTGTTCTTACTACTAATG	3047
Qy	2401	AAGAAACCATGCTTGGAGGGGCGTGAACACAGAACCCCTCAAGACAAGGATGACAGCT	2460
Db	3048	AAGAAACCATGCTTGGAGGGGCGTGAACACAGAACCCCTCAAGACAAGGATGACAGCT	3107
Qy	2461	GGAGGACACATCTAGCTGCATTCGCACTTCGCACTCTGGGCTCCCGAGACTCTGCTGTGTGAGA	2520
Db	3108	GGAGGACACATCTAGCTGCATTCGCACTTCGCACTCTGGGCTCCCGAGACTCTGCTGTGTGAGA	3167
Qy	2521	AATTAAACCCCTGCTTGTCTTGAGAAAAA	2555
Db	3168	AATTAAACCCCTGCTTGTCTTGAGAAAAA	3202
RESULT 4			
AX714233			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Homo sapiens (human)			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE			
AUTHORS			
1			
Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,			
Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,			
Tanemoto, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and			
Masuho, Y.			
TITLE			
Full-length cDNAs			
JOURNAL			
Patent: EP 1293569-A 917 19-MAR-2003;			
Helix Research Institute (JPL) - Research Association for			
Biotechnology (JP)			
FEATURES			
source			
1..3188			
/organism="Homo sapiens"			
/mol_type="genomic DNA"			
/db_xref="taxon:9606"			
BASE COUNT			
ORIGIN			
725 a 909 G 685 t			
Query Match			
Best Local Similarity 99.4%; Score 2538.8; DB 6; Length 3188;			
Matches 2540; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
Qy	1	AGCAGTCCCACTCTATGACCAGTGGGAGGATGTGATGAAGGATGAAGTGGAGGTGC	60
Db	647	AGCAGTCCCACTCTATGACCAGTGGGAGGATGTGATGAAGGATGAAGTGGAGGTGC	706
Qy	61	TCAACAGTGATGCTGTGCTCCCCAGCCGGGTGTACTGATGCTCTGTCTATCCAGACAG	120
Db	707	TCAACAGTGATGCTGTGCTCCCCAGCCGGGTGTACTGATGCTCTGTCTATCCAGACAG	766
Qy	121	CAGGATACGGGTGCTGCTTCGGTATGAAGCTTTGAAAATGACGCCAGCCATGACTTCT	180
Db	767	CAGGATACGGGTGCTGCTTCGGTATGAAGCTTTGAAAATGACGCCAGCCATGACTTCT	826
Qy	181	GTGCAACCTGGGAACAGTGGATGTCCACCCCATTTGGCTGGTGTGCCATCAACAGCAAGA	240
Db	827	GTGCAACCTGGGAACAGTGGATGTCCACCCCATTTGGCTGGTGTGCCATCAACAGCAAGA	886
Qy	241	TCCTAGTGGCCCGACGGACCATTCATGCCAGTTCACCGACTGGAAGGCTACCTCATGA	300
Db	887	TCCTAGTGGCCCGACGGACCATTCATGCCAGTTCACCGACTGGAAGGCTACCTCATGA	946
Qy	301	AACGGCTGGTGGCTCCAGGACGGTTCCTCGTGGATTTCCATCAAGATGTGTGGAGCA	360
Db	947	AACGGCTGGTGGCTCCAGGACGGTTCCTCGTGGATTTCCATCAAGATGTGTGGAGCA	1006
Qy	361	TGAAGTACCCCTTTAGGCAGGGCATGCGGCTGGAGTGGTGGACAGTCCCAGGTGTCTAC	420


```

RESULT 5
AK056443
LOCUS
DEFINITION Homo sapiens cDNA FLJ1881 fis, clone NT2RP7002829, weakly similar
to Scn-related gene containing four mbt domains.
ACCESSION AK056443
VERSION 1
KEYWORDS oligo capping, fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens

REFERENCE
AUTHORS Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A.,
Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S.,
Nagahari K., Masuho Y., Nagai K. and Isogai T.
NEDO human cDNA sequencing project
1 Unpublished
2 (bases 1 to 3188)
Isogai T., Otsuki T. and Sugiyama T.
Direct Submission
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(B-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- and 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: RAB and
HRI.

FEATURES
source
1. 3188
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP7002829"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/notes="cloning vector: pME18SPL3-mRNA from NT2 neuronal
precursor cells after 5-weeks retinoic acid (RA)
induction."

BASE COUNT 725 a 869 c 909 g 685 t
ORIGIN
Query Match 99.4%; Score 2538.8; DB 9; Length 3188;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2540; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGCAGTCCCACTCTATGACCAAGTGGAGGATGTGATGAAGAGGATGAAGGTGGAGGTGC 60
DB 647 AGCAGTCCCACTCTATGACCAAGTGGAGGATGTGATGAAGAGGATGAAGGTGGAGGTGC 706
QY 61 TCACAGTGTGCTGCTCCAGCGGGTGTACTGCTGCTCTGTCTATCCAGACAG 120
DB 707 TCACAGTGTGCTGCTCCAGCGGGTGTACTGCTGCTCTGTCTATCCAGACAG 766
QY 121 CAGGATTCGGGTGCTCTCGGTATGAAGCTTTGAAATGACGCGACCATGATTTCT 180
DB 767 CAGGATTCGGGTGCTCTCGGTATGAAGCTTTGAAATGACGCGACCATGATTTCT 826
QY 181 GGTGCACTGGGACAGTGTGATGTCACCCCATTTGGTGTGCTGATCAACAGCAAGA 240
DB 827 GGTGCACTGGGACAGTGTGATGTCACCCCATTTGGTGTGCTGATCAACAGCAAGA 896

```


Qy	61	TCACAGTGTGCTGCTCCAGCGGGTGTACTGATCGCTCTGTCTATCCAGACAG	120	Qy	1141	AGTGGGTGAGTGTGAGTCCCGAGACATCTACCCGTCGGTGGTGTGAGTCTCACCGGCT	1200
Dd	695	TCACAGTGTGCTGCTCCAGCGGGTGTACTGATCGCTCTGTCTATCCAGACAG	754	Dd	1775	AGTGGGTGAGTGTGAGTCCCGAGACATCTACCCGTCGGTGGTGTGAGTCTCACCGGCT	1834
Qy	121	CAGGTATCGGTGCTGCTGCTGATGAAAGCTTTGAAATGACGCCAGCCATGATCTT	180	Qy	1201	ACGAGCTCCAGCTTCTGTCGCGG-----	1224
Dd	755	CAGGTATCGGTGCTGCTGCTGATGAAAGCTTTGAAATGACGCCAGCCATGATCTT	814	Dd	1835	ACCAGCTCCAGCTTCTGTCGCGCAGGTAACTACCCAGGTCCCTCTGAGCAGCCCTGG	1894
Qy	181	GGTGCAACCTGGGACAGTGGATGCCACCCATTCGCTGGTGGTGCATCAACAGCAAGA	240	Qy	1225	-----	1224
Dd	815	GGTGCAACCTGGGACAGTGGATGCCACCCATTCGCTGGTGGTGCATCAACAGCAAGA	874	Dd	1895	TAAACAGGTGGGAAGGACAGCTGTTCTCCGGCCCCCTCTCCAGCCCCCGCTCTC	1954
Qy	241	TCCTAGTGCCTCCAGCGACCATCCATGCTCAAGTTTCCCGATGGAAGGCTACCTCATGA	300	Qy	1225	-----CAGAACCGGCCACACCCGCTGAAGCCCAAGAGGCCACAAAGAAAGA	1275
Dd	875	TCCTAGTGCCTCCAGCGACCATCCATGCTCAAGTTTCCCGATGGAAGGCTACCTCATGA	934	Dd	1955	CTCATTTCCAGAACCGGCCACACCCGCTGAAGGCCAAAGAGGCCACAAAGAAAGA	2014
Qy	301	AAAGCTGGTGGCTCCAGACGCTTCCCGTGTATTTCCACATCAAGATGGTGGAGACCA	360	Qy	1276	AACAGTTTGGGAAGAAAGAAAGAAATCCCGCCCTCACTAAGACGCGACCCCTCAGACAGG	1335
Dd	935	AAAGCTGGTGGCTCCAGACGCTTCCCGTGTATTTCCACATCAAGATGGTGGAGACCA	994	Dd	2015	AACAGTTTGGGAAGAAAGAAAGAAATCCCGCCCTCACTAAGACGCGACCCCTCAGACAGG	2074
Qy	361	TGAAGTACCCCTTTAGCAGCGGCATCGGCTGGAAGTGTGGACAAAGTCCAGGTGTAC	420	Qy	1336	GGTCCAAAGAGCCCTCTGTGAGAGCAGACCCCTCAGGCTGCAGGAAGATCTCTCGGAGC	1395
Dd	995	TGAAGTACCCCTTTAGCAGCGGCATCGGCTGGAAGTGTGGACAAAGTCCAGGTGTAC	1054	Dd	2075	GGTCCAAAGAGCCCTCTGTGAGAGCAGACCCCTCAGGCTGCAGGAAGATCTCTCGGAGC	2134
Qy	421	GCATCGCATGGTGTGGTGGACACAGTAAATCGGGGGTCCCTACGGCTCTCTACGAGG	480	Qy	1396	CTGTTCTCTGGCGAGATCAATTGCTGTGCTGTGAGGAAGAGCATCTAGAGCTTGCCTCGC	1455
Dd	1055	GCATCGCATGGTGTGGTGGACACAGTAAATCGGGGGTCCCTACGGCTCTCTACGAGG	1114	Dd	2135	CTGTTCTCTGGCGAGATCAATTGCTGTGCTGTGAGGAAGAGCATCTAGAGCTTGCCTCGC	2194
Qy	481	ATGCTGACGTGACGACGATCTTGGTGCCACATGTGGAGCCCTTGATCCACCCAGTGG	540	Qy	1456	CCGACAAAGCTTCAAGTCCAGAGCTGCTGTCTCGCTCGAGAACATCAAGCAGGAACAG	1515
Dd	1115	ATGCTGACGTGACGACGATCTTGGTGCCACATGTGGAGCCCTTGATCCACCCAGTGG	1174	Dd	2195	CCGACAAAGCTTCAAGTCCAGAGCTGCTGTCTCGCTCGAGAACATCAAGCAGGAACAG	2254
Qy	541	GTGGTCAACGAGTGTGGGCCACAGGCATCAAGATGTCCAGAGAGCGGAAGTGCATGGCC	600	Qy	1516	ACGAGTGAAGCTTCCCTCCCTCCAGCTGGCTTCTAGCTGGAAGCCAGCCAGCGCTTCTC	1575
Dd	1175	GTGGTCAACGAGTGTGGGCCACAGGCATCAAGATGTCCAGAGAGCGGAAGTGCATGGCC	1234	Dd	2255	ACGAGTGAAGCTTCCCTCCCTCCAGCTGGCTTCTAGCTGGAAGCCAGCCAGCGCTTCTC	2314
Qy	601	ATACCCCACTTCCGGAAGATCTACTGTGATGTCGCTTCTTACCTTCAAGAGGTAC	660	Qy	1576	TACCACCAACCACTGCTCCACCTGACTTTGGCTTGGAGACTGATCTCTCTGTGAAA	1635
Dd	1235	ATACCCCACTTCCGGAAGATCTACTGTGATGTCGCTTCTTACCTTCAAGAGGTAC	1294	Dd	2315	TACCACCAACCACTGCTCCACCTGACTTTGGCTTGGAGACTGATCTCTCTGTGAAA	2374
Qy	661	GAGCAGTCTACAGAAAGCGGTGGTTTGAAGAGGATGAAGCTGGAGGCCATTGACC	720	Qy	1636	TTCTGCCCGGTGCTGTGAGGCTGGAAGCTGGAAGCTGCTGCTGGGTCTCTCTGGAGCCG	1695
Dd	1295	GAGCAGTCTACAGAAAGCGGTGGTTTGAAGAGGATGAAGCTGGAGGCCATTGACC	1354	Dd	2375	TTCTGCCCGGTGCTGTGAGGCTGGAAGCTGGAAGCTGCTGCTGGGTCTCTCTGGAGCCG	2434
Qy	721	CCCTGAATCTGGGCAACATCTGCGTGGCACTGTCTGAAGTCTCTCTGATGATACC	780	Qy	1696	CTGTTGCTTCTGCGCTTCCCTGCTGGAAGGCTTATATGACGGGCCGCTGAGGCCCCAG	1755
Dd	1355	CCCTGAATCTGGGCAACATCTGCGTGGCACTGTCTGAAGTCTCTCTGATGATACC	1414	Dd	2435	CTGTTGCTTCTGCGCTTCCCTGCTGGAAGGCTTATATGACGGGCCGCTGAGGCCCCAG	2494
Qy	781	TGATGATCTGTGTGGAGCGGGGCCCTCCACAGATGGCTTGGACTGCTTCTGCTACCATG	840	Qy	1756	AATCTGCTGTGAACACCTTTTCCAGCAGAGTTCCTCAAGAGCTGGAACGCTAGCTGCT	1815
Dd	1415	TGATGATCTGTGTGGAGCGGGGCCCTCCACAGATGGCTTGGACTGCTTCTGCTACCATG	1474	Dd	2495	AATCTGCTGTGAACACCTTTTCCAGCAGAGTTCCTCAAGAGCTGGAACGCTAGCTGCT	2554
Qy	841	CCTCTTCCAGGCCATCTTCCCGGCCACTTCTGTGAGAGAGATGATGATGATGATGATG	900	Qy	1816	GCTCTTCTTAAAGTGGCTTCCCGCCAGCCGACCGCCCTCAGTTGCCAGGGATGGGG	1875
Dd	1475	CCTCTTCCAGGCCATCTTCCCGGCCACTTCTGTGAGAGAGATGATGATGATGATGATG	1534	Dd	2555	GCTCTTCTTAAAGTGGCTTCCCGCCAGCCGACCGCCCTCAGTTGCCAGGGATGGGG	2614
Qy	901	CGCAAAAGGTTATGAGCAGACACTTTCAACTGGGAGAACTACTTGGAGAACACCAAGT	960	Qy	1876	CCACCACTGTCACTGTGGAATAAAGACAGTGAATCTGTCTGCTGCTGGAACGAGTCAATG	1935
Dd	1535	CGCAAAAGGTTATGAGCAGACACTTTCAACTGGGAGAACTACTTGGAGAACACCAAGT	1594	Dd	2615	CCACCACTGTCACTGTGGAATAAAGACAGTGAATCTGTCTGCTGCTGGAACGAGTCAATG	2674
Qy	961	CGAAAGCGCTTCCATCGAGACTCTTTAACTATGATGATGATGATGATGATGATGATG	1020	Qy	1936	TAAATTAAGTCTAGAGCAGCTCTCTGAGCAGGATTAAGTCCCTGACAGTGAATGATG	1995
Dd	1595	CGAAAGCGCTTCCATCGAGACTCTTTAACTATGATGATGATGATGATGATGATGATG	1654	Dd	2675	TAAATTAAGTCTAGAGCAGCTCTCTGAGCAGGATTAAGTCCCTGACAGTGAATGATG	2734
Qy	1021	GCATGAAGCTGGAGCGCTGATGATGAGCGCCCGGCTCATCTGTGTGCGCCACGGTGA	1080	Qy	1996	GGTGGGGCAGCCTCTGCTCTCAAAATTCACCAAGCAGAAATGCTCTCAGCCTCATGPT	2055
Dd	1655	GCATGAAGCTGGAGCGCTGATGATGAGCGCCCGGCTCATCTGTGTGCGCCACGGTGA	1714	Dd	2735	GGTGGGGCAGCCTCTGCTCTCAAAATTCACCAAGCAGAAATGCTCTCAGCCTCATGPT	2794
Qy	1081	AACGAGTGGTGCATCGGCTCTCTAGCATCCACTTTGACGGCTGGGACAGCGAGTACGACC	1140	Qy	2056	TGGTCTCTGCTCTCTCTAGCTCCCGAGGATGTGGGACCGAGCTTGTCTCGGAGCT	2115
Dd	1715	AACGAGTGGTGCATCGGCTCTCTAGCATCCACTTTGACGGCTGGGACAGCGAGTACGACC	1774	Dd	2795	TGGTCTCTGCTCTCTCTAGCTCCCGAGGATGTGGGACCGAGCTTGTCTCGGAGCT	2854
				Qy	2116	AAGAGCAGTGAACAGGATGTGGATTTTGGGCGACTGTGTGTGGCTTGGAGCTGCTTTC	2175


```

Db      5796 AGGCACATCTAGTCGCATGACCACTCACTGGCTCCCGACACTCTGTGTGAGAA 5855
      2523 TTAACCCCTGCTGCTTG 2542
      5856 TTAACCCCTGCTGCTTG 5875

RESULT 8
HSM801538      3071 bp      mRNA      linear      PRI 20-MAR-2002
LOCUS      Homo sapiens mRNA; cDNA DKFZp7611141 (from clone DKFZp7611141);
DEFINITION      complete cds.
ACCESSION      AL136564
VERSION      AL136564.1 GI:13276634
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 3071)
AUTHORS      Ottenwaelder, B., Obermaier, B., Mewes, H.W., Gassenhuber, J. and
      Wiemann, S.
TITLE      Direct Submission
JOURNAL      Submitted (12-MAR-2002) MIPS, Am Klopferspitze 18a, D-82152
      Martinsried, GERMANY
COMMENT      Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
      Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
      sequenced by MediGenomix (Martinsried/Germany) within the cDNA
      sequencing consortium of the German Genome Project. This clone
      (DKFZp761141) is available at the RZPD in Berlin. Please contact
      the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
      Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
      information about the clone and the sequencing project is available
      at http://mips.gsf.de/proj/cDNA/.
FEATURES
      Location/Qualifiers.
          1..3071
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /map="22q13.31-13.33"
              /clones="DKFZp7611141"
              /tissue_type="amygdala"
              /clone_lib="761 (synonym: hamy2). Vector pSport1; host
              DH10B; Sites NotI + SalI"
              /dev_stage="adult"
          1..3071
              /gene="DKFZp7611141"
              /db_xref="GeneID:13276634"
              /translation="MEKPSIEETPSSEPMEEEDDDLELFGYDSFRS VNSVSGES
              SSYLEESSEANEDEARELPTSPHLLSPPTPSLDSGSEFAVEMCSIVGTREAF
              FSKTRCSVSCSRSSNSKKASILARLQKPTPKAKVHLKAAWSAKIGAFHSGG
              TQLADGPTQDQALVGFDMGKFLDHSYKAAFPVSCFKEVPLDQMEDYMGKMYEV
              LNSDAVLPSRVYASVQTAGYRVLRYEGFENDASHDFWNCNLGTVDVHPVIGWCAN
              SKILVPRPTIHAETDKNGYLMKRLVGRSLPVDPHI KMVESMKYPPROGMRLVVDK
              SOVSRVAVVDVTVIGRLRLLYEDGSDDDFWCHMKSPHLHPVGMSSRVYGHGKMS
              RSRMAHPTFRKYLCAVPLPKKRVAVTEGWNFEQGMKLEADPLNLGNICVAVI
              CKVLIDGLYMLICVGGSTGDLDFWCFHASSHAI FPAIFCQKNDIEUTPPGVEAQT
              NWENYLEKTSKAPSLFNMDPNHGFVKGMLEAVDLMEPLICVATVYKRVVHRL
              SIHFDGMDSEVDQWDCESDPIYPVGMCELTGLOLPPVAAGVSGSRPKRL"
              /gene="DKFZp7611141"
              /polya_site
          665 a      882 c      958 g      665 t
BASE COUNT
ORIGIN
Query Match      56.1%; Score 1434.6; DB 9; Length 3071;

```

```

Best Local Similarity 79.8%; Pred. No. 0;
Matches 1330; Conservative 0; Mismatches 9; Indels 480; Gaps 1;

QY      1 AGCAGTCCCACTCTATGACACAGTGGGAGGATGATGATGAAGGATGAAGTGGAGTGC 60
      650 AGCAGTCCCACTCTATGACACAGTGGGAGGATGATGATGAAGGATGAAGTGGAGTGC 709
      61 TCACAGTGTATGCTGTGCTCCCGAGCGGGGTACTGGATCGCCTCTGTCTATCAGACAG 120
      710 TCACAGTGTATGCTGTGCTCCCGAGCGGGGTACTGGATCGCCTCTGTCTATCAGACAG 769
      121 CAGGATATCGGGTGTCTGCTCGGTATGAAGCTTTGAAATGACCCAGCCATGACTTCT 180
      770 CAGGATATCGGGTGTCTGCTCGGTATGAAGCTTTGAAATGACCCAGCCATGACTTCT 829
      181 GGTCAACCTCGGACAGTGTATGCCACCCATTTGGTGTGTGCTCATCAACAGCAAGA 240
      830 GGTCAACCTCGGACAGTGTATGCCACCCATTTGGTGTGTGCTCATCAACAGCAAGA 889
      241 TCCTAGTCCCGCCAGCGACCATCCATGCCAAGTTCCAGCTGGAAGGGTACCTCATGA 300
      890 TCCTAGTCCCGCCAGCGACCATCCATGCCAAGTTCCAGCTGGAAGGGTACCTCATGA 949
      301 AACGCTGTGGGCTCCAGGACGCTTCCGCTGGATTCCACATCAAGATGCTGGAGACA 360
      950 AACGCTGTGGGCTCCAGGACGCTTCCGCTGGATTCCACATCAAGATGCTGGAGACA 1009
      361 TGAAGTACCCCTTTAGGACAGGCGATGCGGCTGGAGTGGTGGACAAAGTCCAGAGTGC 420
      1010 TGAAGTACCCCTTTAGGACAGGCGATGCGGCTGGAGTGGTGGACAAAGTCCAGAGTGC 1069
      421 GCATCCGATGCTGTGTGACACAGTAATCGGGGTGCGCTACGGCTCTCTACGAGG 480
      1070 GCATCCGATGCTGTGTGACACAGTAATCGGGGTGCGCTACGGCTCTCTACGAGG 1129
      481 ATGTGTGACAGTGAAGACGACATTTCTGTGCCACATGTGTGAGCCCCCTGATCCACCCAGTGG 540
      1130 ATGTGTGACAGTGAAGACGACATTTCTGTGCCACATGTGTGAGCCCCCTGATCCACCCAGTGG 1189
      541 GTTGTGACAGTGTGGGCGACGCGCATCAAGATGTGAGAGGCGAAGTGAAGTGCATGGGCC 600
      1190 GTTGTGACAGTGTGGGCGACGCGCATCAAGATGTGAGAGGCGAAGTGAAGTGCATGGGCC 1249
      601 ATCAACCCCACTTCCGGAAGTCTACTGTGATCGCGTTCTTACCTCTTCAAGAAGGTAC 660
      1250 ATCAACCCCACTTCCGGAAGTCTACTGTGATCGCGTTCTTACCTCTTCAAGAAGGTAC 1309
      661 GAGCAGTCTTACACAGAGCGGTTGGTTTGAAGAGGAGTGAAGTGAAGGCGCATTGACC 720
      1310 GAGCAGTCTTACACAGAGCGGTTGGTTTGAAGAGGAGTGAAGTGAAGGCGCATTGACC 1369
      721 CCTGAATCTGGGCAACATCTCGGTGGCACTGTCTGTAAAGTTCTCTGGATGGATACC 780
      1370 CCTGAATCTGGGCAACATCTCGGTGGCACTGTCTGTAAAGTTCTCTGGATGGATACC 1429
      781 TGATGATCTGTGTGACGCGGGGCGCTCCACAGATGGGTGGAGTGGTTCTGCTACCATG 840
      1430 TGATGATCTGTGTGACGCGGGGCGCTCCACAGATGGGTGGAGTGGTTCTGCTACCATG 1489
      841 CTTCTTCCACGCCCATCTTCCCGGCACTTCTGTGACAGAAATGACATTTAGCTCACAC 900
      1490 CTTCTTCCACGCCCATCTTCCCGGCACTTCTGTGACAGAAATGACATTTAGCTCACAC 1549
      901 CGCCAAAAGTTATGAGGCACAGACTTTTCACTGGGAGAACTACTTGGAGAGAGACCAAGT 960
      1550 CGCCAAAAGTTATGAGGCACAGACTTTTCACTGGGAGAACTACTTGGAGAGAGACCAAGT 1609
      961 CGAAAAGCCGCTCCATCGAGACTCTTTAAATGATGGATGGCCAAAACCATGGCTTCAAGGTGG 1020
      1610 CGAAAAGCCGCTCCATCGAGACTCTTTAAATGATGGATGGCCAAAACCATGGCTTCAAGGTGG 1669
      1021 GCATGAGTGTGAGGCGCGCTGAGACCTGATGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1080

```

Db 1670 GCATGAAGCTGGAGCGCTGGAACTGATGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1729
Qy 1081 AACGAGTGGTCATCGCTCTCAGCATCCACTTTGACGCTGGGACACGAGTAGACGCC 1140
Db 1730 AACGAGTGGTCATCGCTCTCAGCATCCACTTTGACGCTGGGACACGAGTAGACGCC 1789
Qy 1141 AGTGGTGGACTCGGAGTCCCAAGACATCTACCCCTCGGCTGGTGTGAGCTCACCGGCT 1200
Db 1790 AGTGGTGGACTCGGAGTCCCAAGACATCTACCCCTCGGCTGGTGTGAGCTCACCGGCT 1849
Qy 1201 ACCAGCTCCAGCCTCTGTGGCCG----- 1224
Db 1850 ACCAGCTCCAGCCTCTGTGGCCGAGGTGTGGGCTCTCGTGGCCCTCAAGAGGCTCTGAC 1909
Qy 1225 ----- 1224
Db 1910 TTTCTTTCCCTCTTTTCTCTCTCCCGCCCTGTGCCCATCTCCGTTCTTTGGC 1969
Qy 1225 ----- 1224
Db 1970 ATGAGTGGAGATGTCTCATGAGACCATTAAAGTAGAGAGTGAAGCCCTCACCGAGCC 2029
Qy 1225 ----- 1224
Db 2030 CTGCTCTGACTTCTCTGTCTCCCTTTCCCTCTGGCTCGAGAGCTCTTCTTCATCTT 2089
Qy 1225 ----- 1224
Db 2090 GCCCACTCTGTCATATTTCTGTGCTTGTGACCCAGGTAAACTACCCAGTCCCTCTG 2149
Qy 1225 ----- 1224
Db 2150 AGCAGCCTGTAAACAAGGTGGGAAGAGGACAGCTGTTCTCCGGCCCTCTCCAGC 2209
Qy 1225 ----- 1224
Db 2210 CCGCCCTCTCCTCATTTGCCAGGTTTGGCTTCTGCTTGGGTGTCTGCTGGAGG 2269
Qy 1225 ----- 1224
Db 2270 GTGATGGGTCTCGGATGCGCTGTGCTGTGCTTCCCTCCAGGAGCCCTTCTTCATC 2329
Qy 1225 -----CAGAACCGCCACACCGCTGAGAGGCCAAGAGGCCA 1260
Db 2330 TCTTTCACCTTCTTCAACAACAGAACCGCCACACCGCTGAGAGGCCAAGAGGCCA 2389
Qy 1261 CAAGAGAGAAAGAACAGCTTTGGGAAGAAAGAAAGATCCCGCCCACTAAGACGC 1320
Db 2390 CAAGAGAGAAAGAACAGCTTTGGGAAGAAAGAAAGAAATCCCGCCCACTAAGACGC 2449
Qy 1321 GACCCCTCAGACAGGGGTCAAGAAGCCCTGTGAGGACGACCTCAGGGTGCCAGGA 1380
Db 2450 GACCCCTCAGACAGGGGTCAAGAAGCCCTGTGAGGACGACCTCAGGGTGCCAGGA 2509
Qy 1381 AGATCTCGTCGGAGCTTTCTCGGAGATCATCTGTGCTGTGCTGGAAGGAGAGCATC 1440
Db 2510 AGATCTCGTCGGAGCTTTCTCGGAGATCATCTGTGCTGTGCTGGAAGGAGAGCATC 2569
Qy 1441 TAGACGTGGCTGCGCCGCAAGAGCTTCAAGTCCAGAGCTGCTGTCTCGTCCGAGAAC 1500
Db 2570 TAGACGTGGCTGCGCCGCAAGAGCTTCAAGTCCAGAGCTGCTGTCTCGTCCGAGAAC 2629
Qy 1501 TCAAGCAGGAAACAGACGACTGAGCCTTCTGCTCTCAGCCTGGCTTTAGTGTGAAGCC 1560
Db 2630 TCAAGCAGGAAACAGACGACTGAGCCTTCTGCTCTCAGCCTGGCTTTAGTGTGAAGCC 2689
Qy 1561 AGCCAGCGTTTCTTACCAACACACGATCGCTCCACCTGACTTTGGCTTGGAGACTGA 1620
Db 2690 AGCCAGCGTTTCTTACCAACACACGATCGCTCCACCTGACTTTGGCTTGGAGACTGA 2749
Qy 1621 TCCTCTGTGTAAATTCCTCCCGGTGCTGTGAAGCGTGAAGCGTGGAGGACTGCTGGG 1680
Db 2750 TCCTCTGTGTAAATTCCTCCCGGTGCTGTGAAGCGTGAAGCGTGGAGGACTGCTGGG 2809

Qy 1681 GTCTCTGGGACCGCCTGTTGCTTCTGCTCCCTGCTGAAAAGGTCTATATACGGGC 1740
Db 2810 GTCTCTGGGACCGCCTGTTGCTTCTGCTCCCTGCTGAAAAGGTCTATATACGGGC 2869
Qy 1741 CGCTGAGGCCCCAAGAACTGCTGTGGAACCACTTTTCCAGCCAGAGTTCCCAAGCTG 1800
Db 2870 CGCTGAGGCCCCAAGAACTGCTGTGGAACCACTTTTCCAGCCAGAGTTCCCAAGCTG 2929
Qy 1801 GAACGCTAGCTGCTGCTCTTCTTAAAGTGGCTCCCGCCAGCCCGCCACGCGCTCAG 1860
Db 2930 GAACGCTAGCTGCTGCTCTTCTTAAAGTGGCTCCCGCCAGCCCGCCACGCGCTCAG 2989
Qy 1861 TTGCCAGGATGGGCCCCACCACTGTCTGGAATAAAGACAGTGAACCTGTGCTG 1920
Db 2990 TTGCCAGGATGGGCCCCACCACTGTCTGGAATAAAGACAGTGAACCTGTGCTG 3049
Qy 1921 CCTGAACGAGTCATGTAAA 1939
Db 3050 CCTAAAAAATAAAAAA 3068
RESULT 9
AK097052
LOCUS Homo sapiens cDNA FLJ39733 fis, clone SMINT2016122.
DEFINITION AK097052
ACCESSION AK097052
VERSION AK097052.1 GI:21756694
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
Kawai H., Saito K., Nishikawa T., Kimura K., Yamashita H.,
Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
Wagatsuma M., Murakawa K., Kanemori K., Takahashi-Fujii A.,
Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S.,
Nagahari K., Maehuo Y., Nagai K. and Isogai T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2432)
Isogai T. and Yamamoto J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 252-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
source
1..2432
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="SMINT2016122"
/tissue_type="small intestine"
/clone_lib="SMINT2"
/note="Cloning vector: pME18SFL3"
397..1977
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC04936.1"
/db_xref="GI:21756695"
/translation="MGIIVTREAFFSKTRFCVSCSRYSNKKASILARIQKRP"
CDS

PTKAKVHLKAAWSAKIGAFHLSQGTQOLADGTPGTQDALVGLDNGKFLKDRHSYKAA
PVSCFKHVLVDQEDVNMKMEVLSNDVLPSTVYIASVITAGYVLLVYEGFE
NDASHDFWNLGTVDVHPHIGWCAINSKILVPTPHIAKFTDMKGYLMKRLVGRKTLFP
DFHUKWDSMKYPRQGRWLEVDKSVSRTRMAVVDVTVIGRLRLLYEDGSDDDFW
CHMSPLIHPVMSRVRGHGIKMSERSDMAHPTFKIYCDVAPVLPKVRVAVYEG
GKFEQMKLEALDPLNLGNICVATUCKVLDGILMICVDGSPSTDGLDWFCHASSHA
IFPAIFCKNDLIEUTPPKGYEAFINWENILEKTSKAAAPSRLFNMDCPNHFVKGMK
LEADFMKPRICVATKRVVHRLLSLHFDGWDSYDQWVDCSPDIPYVGVGCELTGY
QLQPPVAGVSGRGPKL"

BASE COUNT 504 a 682 c 696 g 550 t
ORIGIN

Query Match 48.1%; Score 1227.8; DB 9; Length 2432;
Best Local Similarity 99.4%; Pred. No. 1.1e-303;
Matches 1232; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 AGCAGTCCCACTCTATGACCAAGTGGGAGGATGTGATGAAAGGATGAAGTGGAGGTGC 60
DB 719 AGCAGTCCCACTCTATGACCAAGTGGGAGGATGTGATGAAAGGATGAAGTGGAGGTGC 778
QY 61 TCAACAGTGTCTGTGTCCTCCAGCGGGGTGTACTGATCGCTCTGTGATCCAGACAG 120
DB 779 TCAACAGTGTCTGTGTCCTCCAGCGGGGTGTACTGATCGCTCTGTGATCCAGACAG 838
QY 121 CAGGATATCGGTGCTGCTCGGTATGAAGCTTTGAAATGACGCGAGCCATGACTTCT 180
DB 839 CAGGATATCGGTGCTGCTCGGTATGAAGCTTTGAAATGACGCGAGCCATGACTTCT 898
QY 181 GGTCAACTGGGAACAGTGGATGTCCACCCCATTTGGTGGTCCATCAACAGCAAGA 240
DB 899 GGTCAACTGGGAACAGTGGATGTCCACCCCATTTGGTGGTCCATCAACAGCAAGA 958
QY 241 TCCTAGTCCCGCCAGGACATCATGCGAGTTCAGCGAGTGAAGGCTACCTCATGA 300
DB 959 TCCTAGTCCCGCCAGGACATCATGCGAGTTCAGCGAGTGAAGGCTACCTCATGA 1018
QY 301 AACGGCTGGTGGGTCCAGGACGCTTCCCGTGAATTTCCATCAAGATGGTGGAGACA 360
DB 1019 AACGGCTGGTGGGTCCAGGACGCTTCCCGTGAATTTCCATCAAGATGGTGGAGACA 1078
QY 361 TGAAGTACCCCTTTAGGAGGCGATGGGCTGGAAGTGGTGGACAAATCCAGGTGTAC 420
DB 1079 TGAAGTACCCCTTTAGGAGGCGATGGGCTGGAAGTGGTGGACAAATCCAGGTGTAC 1138
QY 421 GCATCTCGATGGCTGTGTGGACACAGTAATCGGGGTGCGCTACGCTCTCTACGAGG 480
DB 1139 GCATCTCGATGGCTGTGTGGACACAGTAATCGGGGTGCGCTACGCTCTCTACGAGG 1198
QY 481 ATGTGACAGTGCAGCAGCTTCTGTGTCACATGTGGAGCCCTGATCCACCCAGTGG 540
DB 1199 ATGTGACAGTGCAGCAGCTTCTGTGTCACATGTGGAGCCCTGATCCACCCAGTGG 1258
QY 541 GTTGTGACGAGCTGTGGGACCGCATCAGATGTGAGAGCGGAGTGCATGAGCC 600
DB 1259 GTTGTGACGAGCTGTGGGACCGCATCAGATGTGAGAGCGGAGTGCATGAGCC 1318
QY 601 ATCAACCCACCTTCGGGAAGATCTACTGTGATGCGCTTCTTACTCTTCAAGAGGTAC 660
DB 1319 ATCAACCCACCTTCGGGAAGATCTACTGTGATGCGCTTCTTACTCTTCAAGAGGTAC 1378
QY 661 GAGCAGTCTACAGAGCGCGTGTGTGTGAGGAAGGATGAGCTGGAGGCGCATGACC 720
DB 1379 GAGCAGTCTACAGAGCGCGTGTGTGTGAGGAAGGATGAGCTGGAGGCGCATGACC 1438
QY 721 CCCTGAATCTGGGAACATCTGCTGGCACTGTCTGAAGTGTCTCTGATGGATACC 780
DB 1439 CCCTGAATCTGGGAACATCTGCTGGCACTGTCTGAAGTGTCTCTGATGGATACC 1498
QY 781 TGATGATCTGTGAGAGCGGGGCGCTCCACAGATGCTTGGAGTGTCTGTACCATG 840
DB 1499 TGATGATCTGTGAGAGCGGGGCGCTCCACAGATGCTTGGAGTGTCTGTACCATG 1558
QY 841 CCTCTTCCAGCCATCTTCCCGGCCACCTTCTGTCAGAGATGATGACTGAGCTACAC 900

1559 CCTCTTCCACGCCATCTTCCCGGCCACCTTCTGTGTCAGAGATGACATGAGTCACAC 1618
QY 901 CGCCAAAAGTTATGAGGCACAGACTTTCACTGGGAGAACTACTTGGAGAGACCAACT 960
DB 1619 CGCCAAAAGTTATGAGGCACAGACTTTCACTGGGAGAACTACTTGGAGAGACCAACT 1678
QY 961 CGAAAGCCCTCCATCGAGACTCTTTAAACATGATGTCGCAACCACTGCTTCAAGGTGG 1020
DB 1679 CGAAAGCCCTCCATCGAGACTCTTTAAACATGATGTCGCAACCACTGCTTCAAGGTGG 1738
QY 1021 GCATGAGCTGGAGCGCTGGACCTGATGGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1080
DB 1739 GCATGAGCTGGAGCGCTGGACCTGATGGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1798
QY 1081 AACGAGTGTGATCGGCTCTTCACTGATCCACTTTTCAAGGCTGGGACGAGTACGACC 1140
DB 1799 AACGAGTGTGATCGGCTCTTCACTGATCCACTTTTCAAGGCTGGGACGAGTACGACC 1858
QY 1141 AGTGGTGTGACTGGAGTCCCGACAGATCTACCCGCTGGCTGTGTGAGCTCAGCGCT 1200
DB 1859 AGTGGTGTGACTGGAGTCCCGACAGATCTACCCGCTGGCTGTGTGAGCTCAGCGCT 1918
QY 1201 ACCAGTCCAGCTCTCTGTGGCGGACGAAACCGGCCACAC 1239
DB 1919 ACAGCTCCAGCTCTCTGTGGCGGACGAGTGTGGGCTCTC 1957

RESULT 10
BC030864
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

BC030864 3395 bp mRNA linear ROD 16-APR-2003
Mus musculus RIKEN cDNA 4732493N06 gene, mRNA (cDNA clone MGC:31247 IMAGE:4211489), complete cds.
BC030864
BC030864.1 GI:21410117
MGC.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 3395)
Strausberg, R.D., Feingold, E.A., Grouse, L.H., Derge, J.G., Schuler, G.D., Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schaefer, C.F., Bhat, N.K., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Wang, J., Wang, J., Hsieh, F., Hopkins, R.F., Jordan, H., Moore, T., Max, J., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stabile, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Winkler, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Kzywinski, M.I., Skalska, U., Smal, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT

human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
2338257
12477932
2 (bases 1 to 3395)
Strausberg, R.
Direct Submission
Submitted (03-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk

QY 1321 GACCCCTCAGACAGGGTCCAGAGAGCCCTGCTGGAGGACGACCCCTCAGGTCGCCAGGA 1380
 Db 1931 GGCCCTCAGACAGGGTCCAGAGAGCCCTTACTGGAGGACAACCTTGAGGCTTTGGG- 1989
 QY 1381 AGATCTCGTGGAGCCTGCTCCGCGAGATCATTTGCTGCGTGTGCAAGGAGACATC 1440
 Db 1990 -----GTCCTGGAACCACTGCTTCCTGATGACATTTGCTGTGTGTGGAAGGAGGACCC 2044
 QY 1441 TAGACGTGGCTCCGCCGACAGAGCTTCAAGTCCAGAGTCCCTGCTGCTCCGTCGAGAAC 1500
 Db 2045 AGGACATTTCTGCTGCTGACAGGTCGCCAGTCCACAGCTGCTCTCCCATTTGAGGCA 2104
 QY 1501 TCAAGCAGGAAACAGACATGAGCCTT-CCTGCTCCAGCCTGGCTTCTAGCTGGAAGC 1559
 Db 2105 TCAAGCAGGAGAGGAAACAACTGAGACTTCCCTGGCATCAGCCTGGACCCCTAACTGAAGCC 2164
 QY 1560 CAGC 1563
 Db 2165 AAGC 2168

RESULT 11

HS756G23 89948 bp DNA linear PRI 12-JAN-2001
 LOCUS Human DNA sequence from clone RP4-756G23 on chromosome
 DEFINITION 2q13.31-13.33 Contains the 5' part of a gene similar to drosophila
 transcriptional repressor, the 3' end of the gene for a novel
 Leucine Rich Protein, the RANGAP1 gene for Ran GTPase activating
 protein 1, ESTs, STS, GSSs and three putative CpG islands, complete
 sequence.
 ACCESSION AL035681
 VERSION AL035681.13 GI:4902689
 KEYWORDS HNG; CpG island; Ran GTPase activating protein; RANGAP1.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 89948)

REFERENCE

AUTHORS Direct Submission
 TITLE Submitted (11-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 JOURNAL requests: clones@sanger.ac.uk

COMMENT

On May 27, 1999 this sequence version replaced gi:4775627.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WormPeP; Information
 on the WormPeP database can be found at
<http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence
 has been finished according to sequence map criteria as follows.
 An attempt is made to resolve all sequencing problems, such as
 compressions and repeats, but not necessarily within known
 annotated repeat sequence elements. Where the sequence is
 ambiguous, there is an annotation using the 'unseq' feature key.
 This sequence was generated from part of bacterial clone contigs of
 human chromosome 22, constructed by the Sanger Centre Chromosome 22
 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr22>
 RP4-756G23 is from the library RPCI-4 constructed by the group of
 Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.hcm>
 VECTOR: pCYPAC2
 IMPORTANT: This sequence is not the entire insert of clone
 RP4-756G23 It may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true left end of clone RP4-756G23 is at 1 in this sequence. The

true left end of clone RP5-979N1 is at 89849 in this sequence. The
 true right end of clone RP1-85F18 is at 80441 in this sequence.

FEATURES

Location/Qualifiers
 1. 89948
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="22"
 /map="ql3.31-13.33"
 /clone="RP4-756G23"
 /clone_lib="RPCI-4"
 2172..2465
 /note="Alusgl repeat: matches 1..296 of consensus"
 complement(10832..11352)
 /note="match: GSS: Em:AQ266638"
 complement(10989..11290)
 /note="match: GSS: Em:AQ269066"
 11441..11790
 /note="match: GSS: Em:AQ204186"
 11455..11594
 /note="match: GSS: Em:AQ285637"
 17446..24698
 /gene="dJ756G23.3"
 join(17446..17677,20088..20170,20515..20749,21230..21295,
 22966..23082,23565..24698)
 /gene="dJ756G23.3" (novel protein similar to drosophila
 transcriptional repressor)
 /note="match: ESTs: Em:H22704 Em:AA351643 Em:AI364341
 Em:T995339 Em:R06448 Em:R36553 Em:R17171 Em:D76919
 Em:AA066436 Em:A459129 Em:AI660728 Em:AA719364
 Em:AI361918 Em:A587146 Em:AI150426 Em:AI160979
 Em:AI522858 Em:A537859 Em:AA281766 Em:AA233695 Em:R27062
 Em:AA905213 Em:R32930 Em:AA356192 Em:AA026761 Em:AA353188
 Em:AA280121 Em:NS6014 Em:TS8551"
 /evidence=not experimental
 CDS
 join(17446..17677,20088..20170,20515..20749,21230..21295,
 22966..23082,23565..23677)
 /gene="dJ756G23.3"
 /note="match: proteins: Tr:O75996 Tr:Q24191 Wp:CE06250
 Tr:O76931"
 /codon_start=1
 /evidence=not experimental
 /product="dJ756G23.3 (novel protein similar to drosophila
 transcriptional repressor)"
 /protein_id="CAB63071.1"
 /db_xref="GI:6572271"
 /db_xref="SPTREMBL:Q9UGS4"
 /translation="MVESMKYPRQGMRLVVDKSVSRTRMAVVDVIGRLRLIYE
 DGSDDDFWCHMSPLIHPVGSRRVGHGIMSGYEAQTENWENYLEKSKAAPSRL
 FNMDPNHGFVKMGKLEAVDLMPRLICVATKRVVHRLLSIHPDGWDSYDQWDCR
 SPDIYPVGMCELTGYQLQPPVAEPATPLKATKKKKQFGKKRIIPPTKPLR
 QGSKKPLEDDPQARKISSEPVPGHIIAVRVKEHLVDVAPDKASPPELPVSVENIK
 QETDD"
 complement(24311..24697)
 /note="match: STS: Em:G14475"
 complement(24350..24620)
 /note="match: STS: Em:G43605"
 complement(24449..24697)
 /note="match: STS: Em:G14931"
 24698
 /gene="dJ756G23.3"
 complement(join(28566..28769,29910..30076,30602..32311,
 32869..33046,34152..34282))
 /gene="dJ756G23.1"
 complement(join(28566..28769,29910..30076,30602..32311,
 32869..33046,34152..34282))
 /note="match: CDNAS: Em:U08018
 match: ESTs: Em:AA128077 Em:AA326017 Em:W58996 Em:AI460232
 Em:AA367412 Em:AA126901
 match: proteins: Tr:O55226 Sw:O02833 Sw:P22792 Sw:Q01129
 Tr:O86637"

```
/codon_start=1
/evidence=not_experimental
/product="dj756G23.1 (novel Leucine Rich Protein)"
/protein_id="CAB63072.1"
/db_xref="GI:6572272"
/db_xref="SPTREMBL:Q9UGS3"
/translation="MVQGIQSSVAVASQKSLPPLGLSLRAGGQGLMGLPVLQEGGLPR
SRTHVPLVLLVLLAPARAAQCPQACDNRHRAVACRYQNLTEVDAIP
EITQRDLQGNLLKVPFAAFQGVFHLTHLDHCEVBLVAEAGFRGLGRILLNLAS
NHLRLEPOALDGLSLRDLLEGNALBELRPGTFGALCALATNLNAHNLVYLPA
FOGLRVRLRSHNALSVALEALGAPALRRLSHNELQALPGPVLQARGLARL
ELGHNPITYAGREDDGALPGLRELLDGCALGALFRAPAFCHLRLDLRGNQDLT
PFGQEQQRLLRWCQCARPILLEWLRARVDSQACQFTHLRGRGADALR
PWLRCFGAAQEEELERAVAGRPFPFPGFERAVCPRAVCVPSRHS
SEGGCGLQVAPGSDTQLDLRRNHPSPVPAFGLHVLVSLHLCQIALEAG
ALAGRLYLYLSNQLAGSAALEGAPRGLYGLFERNRFLQVFGAALRALPSLFS
LHLQDNVRLAPDGLRFRALRVVYLSGNRTYVSLGALGPARELKHLDNRQLE
VPTGALGEPALLELQLSNPLRALRDGAQFQVGHSLQHLPLNSSLGLQICPGAFSG
FGQLQSHLQKQLRALPALPSLSQLELIDLSNPFHCDQLPLRHLWLTGLNLRVGA
TCATPPNAGQRVAAAFAFEDPCGWAARKAKRTFASRPTPIKRGQCAKRVG"
30943..32264
/notes="CpG island"
/evidence=not_experimental
/complement(32183..32620)
/gene="dj756G23.1"
/notes="match: GSS: Em:AQ389159"
32627..33079
/notes="match: GSS: Em:AQ383531"
33161..33287
/notes="AluJo/FLAM repeat: matches 1..131 of consensus"
33563..35664
/notes="53 copies 2 mer ta 86% conserved"
36802..36970
/notes="MER97a repeat: matches 1..178 of consensus"
complement(39025..79639)
/gene="RANGAP1"
complement(join<35028..40098,42756..42877,43155..43243,
44433..44535,46298..46417,47734..47920,49447..49531,
49617..49716,50137..50350,51374..51532,54872..55006,
58090..58269,61523..61582,68026..68153,74359..74555,
79412..79639))
/gene="RANGAP1"
/product="dj756G23.2 (Ran GTPase activating protein 1)"
/notes="match: CDNAs: Em:U08110 Em:X82260 Em:U20857
Em:U88155 Em:AB02306 Em:D79984 Em:X07648
match: ESTs: Em:AA835922 Em:AA280707 Em:AA351030
Em:AA789231 Em:AA985159 Em:H46148 Em:T08060 Em:H21130
Em:T08061 Em:AA026670 Em:AA412337 Em:T35912 Em:D82175
Em:A1139769 Em:A1203851 Em:AA852745 Em:AA412447
Em:AA026631 Em:R26090 Em:H17609 Em:H47302 Em:AA991855
Em:T30135 Em:N90379 Em:AA769102 Em:A1138957 Em:AA852746
Em:AA552593 Em:H18108 Em:H18150 Em:W52172 Em:AA324880
Em:R78185 Em:H68027 Em:AA343282 Em:AA383824 Em:H40643
Em:H50021 Em:AA371797 Em:H55068 Em:D25677"
/evidence=not_experimental
/complement(39025)
/gene="RANGAP1"
39047..39317
/notes="match: STS: Em:G20672"
complement(39054..39059)
/gene="RANGAP1"
complement(39301..39445)
/notes="match: STS: Em:G44270"
complement(join(40029..40098,42756..42877,43155..43243,
44433..44535,46298..46417,47734..47920,49447..49531,
49617..49716,50137..50250,51374..51532,54872..55006,
58090..58269,61523..61582,68026..68153,74359..74470))
/gene="RANGAP1"
/notes="match: proteins: Sw:P46060 Sw:P46061 Sw:O13066
Tr:O17437 Tr:O36421"
/codon_start=1
/evidence=not_experimental
/product="dj756G23.2 (Ran GTPase activating protein 1)"
/protein_id="CAB63073.1"
/db_xref="GI:6572273"
/db_xref="SWISS-PROT:P46060"
/translation="MASEIDIAKLAETLAKTVAGGQLSPKGSKLKLTAEADAKDVIKE
Query Match 44.3%; Score 1132.4; DB 9; Length 89948;
Best Local Similarity 99.9%; Pred. No. 3.3e-279;
Matches 1133; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1411 TCATTGCTGTCGCTGTGAAGGAAGAGCATCTAGACGTGGCTCGCCGACCAAGCGCTTCAA 1470
DB 23565 TCATTGCTGTCGCTGTGAAGGAAGAGCATCTAGACGTGGCTCGCCGACCAAGCGCTTCAA 23624
QY 1471 GTCAGAGCTGCCCTGTCTCCGTGAGAACATCAAGCAGGAGAAACAGACGACTGAGCCTTCC 1530
DB 23625 GTCAGAGCTGCCCTGTCTCCGTGAGAACATCAAGCAGGAGAAACAGACGACTGAGCCTTCC 23684
QY 1531 TGCTCCAGCTGGCTTCTAGCTGGAGCCAGCCAGCGTTTCTTACCAACACCAACCAT 1590
DB 23685 TGCTCCAGCTGGCTTCTAGCTGGAGCCAGCCAGCGTTTCTTACCAACACCAACCAT 23744
QY 1591 GCCTCCACCTGACTTTGGCTTGGAGACTGATCTCTCTGTGTAAATTTCTGCCCGGTGCTG 1650
DB 23745 GCCTCCACCTGACTTTGGCTTGGAGACTGATCTCTCTGTGTAAATTTCTGCCCGGTGCTG 23804
QY 1651 TGAAGGCTGACCGGTGAGGACCTGCTGGGGTCTCTGGGACCGCGCTGTGTTCTTCC 1710
DB 23805 TGAAGGCTGACCGGTGAGGACCTGCTGGGGTCTCTGGGACCGCGCTGTGTTCTTCC 23864
QY 1711 CTCCTCTGTGGAAGGTCTATATGACGGGCGGCTGAGGCCCGCCAGAACTCGTCTGTGAAC 1770
DB 23865 CTCCTCTGTGGAAGGTCTATATGACGGGCGGCTGAGGCCCGCCAGAACTCGTCTGTGAAC 23924
QY 1771 CACTTTTTCAGCCAGAGTTTCCAAAGCTGGAAAGCTAGTGCCTGCTTCTTCTTAAGAT 1830
DB 23925 CACTTTTTCAGCCAGAGTTTCCAAAGCTGGAAAGCTAGTGCCTGCTTCTTCTTAAGAT 23984
QY 1831 GGCTCTCCCGGACCGCCAGCGCTCAGTTCACAGGAGTGGGCGCCACCACTGTGCACAC 1890
DB 23985 GGCTCTCCCGGACCGCCAGCGCTCAGTTCACAGGAGTGGGCGCCACCACTGTGCACAC 24044
QY 1891 TGTGGAAATACAAGACAGTGAACCTGTCTGCCCTGAACAGAGTCATGTAAATTAAGTTCTAG 1950
DB 24045 TGTGGAAATACAAGACAGTGAACCTGTCTGCCCTGAACAGAGTCATGTAAATTAAGTTCTAG 24104
QY 1951 AGCAGCTCTCTGAGCAGGATAAGGTCCTCCCTGACAGTGAAGTGTGTTGGTGGGCGAGCCTC 2010
DB 24105 AGCAGCTCTCTGAGCAGGATAAGGTCCTCCCTGACAGTGAAGTGTGTTGGTGGGCGAGCCTC 24164
QY 2011 TGCCTCAAAAATTCACCAAGCAGAAATGCTCTCTCAGCCTCATGTGTGGTGCCTCTGCTCTCT 2070
DB 24165 TGCCTCAAAAATTCACCAAGCAGAAATGCTCTCTCAGCCTCATGTGTGGTGCCTCTGCTCTCT 24224
QY 2071 CCTAGCTCCCGAGGATGTTGGGAGCCAGCTGTCTCGCAGCTAAGAGCAGTGAACCA 2130
DB 24225 CCTAGCTCCCGAGGATGTTGGGAGCCAGCTGTCTCGCAGCTAAGAGCAGTGAACCA 24284
QY 2131 GGATGTGATTTTGGCGACCTGTGTGGTGGCTTGTAGCTGCTTCTGTGTTTGTGAGGAC 2190
DB 24285 GGATGTGATTTTGGCGACCTGTGTGGTGGCTTGTAGCTGCTTCTGTGTTTGTGAGGAC 24344
QY 2191 TGACTCCCATTTTCTAAAGAAATGCCCCGGGAGGACATTTGGAGGAGAAAGATGGCCCTGA 2250
DB 24345 TGACTCCCATTTTCTAAAGAAATGCCCCGGGAGGACATTTGGAGGAGAAAGATGGCCCTGA 24404
QY 2251 GTGTGCACTTTGGCTCTGCTACCTGCTCCCTGAAGCCCGCTTAAATAATTAATTCATCCAAGA 2310
DB 24405 GTGTGCACTTTGGCTCTGCTACCTGCTCCCTGAAGCCCGCTTAAATAATTAATTCATCCAAGA 24464
QY 2311 TTCCTTTGTAGTTAAAGGTCCTCAAGTTCCTGACTGGAGCTCTTAGAGAGCTGGGCTTGTATG 2370
DB 24465 TTCCTTTGTAGTTAAAGGTCCTCAAGTTCCTGACTGGAGCTCTTAGAGAGCTGGGCTTGTATG 24524
```

```
QY 2371 TTCTTTGGCCTTTTGTCTTACCTAAATGAAGAAACCATGCTGGAGGGCCCTGAACA 2430
Db 24525 TTCTTTGGCCTTTTGTCTTACCTAAATGAAGAAACCATGCTGGAGGGCCCTGAACA 24584
QY 2431 CAGAACCTCCAGACAGGATGACAGCTGGAGGACACATCTAGCTGCAATGCAACCT 2490
Db 24585 CAGAACCTCCAGACAGGATGACAGCTGGAGGACACATCTAGCTGCAATGCAACCT 24644
QY 2491 CACTGGGCTCCCGACAGCTGTGTGTGAGAAATTAACCCCTGCTTCTTTGAG 2544
Db 24645 CACTGGGCTCCCGACAGCTGTGTGTGAGAAATTAACCCCTGCTTCTTTGAG 24698

RESULT 12
AY237001
LOCUS Mus musculus M4MBT variant B (M4mbt) mRNA linear ROD 15-APR-2003
DEFINITION Mus musculus M4MBT variant B (M4mbt) mRNA, complete cds;
alternatively spliced.
ACCESSION AY237001
VERSION AY237001.1 GI:29569825
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 3399)
AUTHORS Markus,J., Feikova,S., Sramko,M., Wolff,L. and Bies,J.
TITLE Cloning and molecular characterization of the novel murine gene
M4mbt encoding a nuclear zinc finger protein with four mbt domains
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3399)
AUTHORS Markus,J., Feikova,S., Sramko,M., Wolff,L. and Bies,J.
TITLE Direct Submission
JOURNAL Submitted (14-FEB-2003) Department of Molecular Virology, Cancer
Research Institute of Slovak Academy of Sciences, Vlnarska 7,
Bratislava, Slovak Republic
FEATURES
source
1..3399
/organism="Mus musculus"
/mol_type="mRNA"
/strain="Swiss Webster"
/db_xref="taxon:10090"
1..3399
/gene="M4mbt"
46..2187
/gene="M4mbt"
/note="alternatively spliced"
/codon_start=1
/product="M4MBT variant B"
/protein_id="AA084917.1"
/db_xref="GI:29569826"
/translation="MEKPRGTBEAPSPSEPMEEEDLDLFGGYDPSRYSNSAGSES
SSYLESEANEDEAREGELPTSTHLPSSANNRSLDGSSEPAVCEMCGIVGTREAP
FSKTRFCVSCRSYSNSKKSLILARLQKPTKKAKVLHKAASAKIGAFLLHAGG
TQCLADPTQDALVDFDNGKFLKDHYSKAAEVSFCFHVPLYDOWEDYMGKMKVEV
LNSADVLPSTVYATVIAQAGYRVLLRYEGFENDASHDFWNLGTVDVHPVIGWCALN
SKILVPPRTIAKTWDKSYLMKRLVGSRTLPAIDPHIKVYESMKYPRQGRLEVVDK
TOVSTRNAVVDTWIGGRLLLYEDGSDDDFWCHMSPLIHPVMSRRVYGHGKMSD
RCDNSHPTREKLYCDAPVYLKRVAVYTEGFWFEGMKLEAIDELNLGSLCYATI
CKVLDDGYLMICDGGPSTDSGDFCYHASSHAIFPATFCQNDIELTPRGYETQPF
SIHFDGWNEDYDQWDESDPIYVQCMLTGYQLQPPVSAEPNTPQKXDTTKKKKK
QFGKKRKIPSAKTRPLRQGSKPLLEDNLLEALGVSEPVDPDSRPVHSCLPPLRASSR
RGTTETSLASAWTLTEAKHGAERREGFTAVQ"
313..387
/misc_feature
/gene="M4mbt"
/note="Region: putative zinc finger"
580..894
/misc_feature
/gene="M4mbt"
/note="Region: mbt repeat"
916..1212
/misc_feature
/gene="M4mbt"
```

```
misc_feature
1267..1545
/notes="Region: mbt repeat"
/genes="M4mbt"
/notes="Region: mbt repeat"
misc_feature
1567..1857
/genes="M4mbt"
/notes="Region: mbt repeat"
BASE COUNT 825 a 905 c 947 g 722 t
ORIGIN
Query Match 44.2%; Score 1128.2; DB 10; Length 3399;
Best Local Similarity 88.9%; Pred. No. 3.9e-278; Indels 0; Gaps 0;
Matches 1220; Conservative 0; Mismatches 153;
QY 1 AGCACATCCACATCTATGACCAAGTGGAGGATGTGATGAAAGGATGAAGGTGGAGTGC 60
Db 641 AACATGTACCCCTCTATGACCAAGTGGAGGAGCTCATGAAGGGGATGAAGGTGGAGTGC 700
QY 61 TCACACAGTGAATGCTGTGCTCCCGAGCGGGTGTACTGGATCGCCTGTGTCATCCACAGAG 120
Db 701 TCAACAGCGATGCTGTGCTCCCGAGCGGGTGTACTGGATCGCCTGTGTCATCCAGGCAG 760
QY 121 CAGGGTATCGGGTGTGCTTCCGTATGAAGGCTTTGAAATGACGCCAGCCATGACTTCT 180
Db 761 CTGGGTACCGGGTGTGCTCCGATATGAAGGCTTTGAAATGACGCCAGCTCATGACTTCT 820
QY 181 GGTGCAACCTGGGAAACAGTGGATGTCCACCCATTGGCTGGTGTGTCATCAACAGCAAGA 240
Db 821 GGTGCAACCTGGGAACTGTGGATGTCCACCCATTGGTGTGTCATCAACAGCAAGA 880
QY 241 TCCTAGTGGCCCCACAGCACCATCCATGCCAAGTTCAACGACTGGAAGGGCTACCTCATGA 300
Db 881 TCCTGTACTCTCACAGCACCATCCATGCCAAGTTTACTGACTGGAGAGACTACCTCATGA 940
QY 301 AACGGCTGTGGGCTCCAGGACGCTTCCGTGGATTCACATCAAGATGGTGGAGAGCA 360
Db 941 AGCGGTGTGGGCTCCAGGACACTTCTCTCGACAGCTTCCATATCAAGATGGTGGAAAGCA 1000
QY 361 TGAAGTACCCCTTTAGCGAGGGCANTCGGCTGGAGTGGTGACAGTCCCGAGGTGTCTAC 420
Db 1001 TGAAGTACCCCTTTCCAGCAGGGGATCGGCTTAGAGTTGTAGCAAGATCAGGTGTCTAC 1060
QY 421 GCATCGCATGGCTGTGTGGACACAGATTAATCGGGGTCGCCCTACGGCTCCCTCTACGAG 480
Db 1061 GTACCCGCATGGCGGTGTGGACACAGATTAATCGGGGTCGCCCTCCGGCTCCCTCTACGAG 1120
QY 481 ATGGTGACAGTACGACGACTTCTGGTGCCACATGTGGAGCCCCCTGTATCCACCCAGTGG 540
Db 1121 ATGGTGACAGTGTATGACGACTTCTGGTGCCATATGTGGAGTCCCTGTATCCACCCAGTGG 1180
QY 541 GTTGGTCAAGAGCTGTGGGCCACGGCATCAAGATGTGAGAGGGGAGGAGTGCATGGCCCC 600
Db 1181 GTTGGTCCGGGTGTGTGGCCACGGCATCAAGATGTGAGAGGGGAGTGCATGGTCTC 1240
QY 601 ATCACCCCACTCCCGAAGATCTACTGTGATGCGCTTCCCTTACCTTTTCAAGAGAGGTAC 660
Db 1241 ATCACCCCACTCCCGAAGATCTACTGTGATGCTGTACCTTTACCTTTTCAAGAGAGGTCC 1300
QY 661 GAGCAGTCTACACAGAGCGGCTTGGTTTGGAGAGGGATGAAGTGGAGGCCATTGACC 720
Db 1301 GCCTGTCTACAGAGGTGGCTGGTTCCGAGAGAGGAGTGAACCTAGAGGCCATTGACC 1360
QY 721 CCCTGAATCTGGGCAACATCTCGGTGGCAACTGTCTGTAAAGTCTTCTCTGTGATGATACC 780
Db 1361 CTCTGAATCTGGGCACTATCTGTGTAGCAACCATCTGCAAGGCTCTCTTGTGATGGTTACC 1420
QY 781 TGATGATCTGTGGACGGGGGGCCCTCCACAGATGGCTTGGAGCTGTCTCTGCTACCATG 840
Db 1421 TGATGATCTGTGGATGGGGGGCCCTCCACAGATGGCTTCCGATGGTCTCTCTACCATG 1480
QY 841 CCTCTTCCACGCCATCTTCCCGGCCACCTTCTGTGAGAGGATGACATTTAGCTGACACAC 900
Db 1481 CCTCTTCCCATGCCATCTTCCCGGCCACCTTCTGCGCAAGAGATGACATTTAGCTGACACAC 1540
```


QY 901 CGCAAAAGGTTATGAGGCACAGACTTTCAACTGGGAGAACTACTTGGAGAACCAAGT 960
 Db 1541 CCCCAAAAGGTTATGAGGCACAGACTTTTGGCTGGGAGAACTACTTAGAGAACCAAGT 1600
 QY 961 CGAAGCGCTCCATCGAGACTCTTTAACATGATTTGCCCAACCATGGCTTCAAGTGG 1020
 Db 1601 CAAGAAGCTGCCAGCAAGACTCTTTAACATGATTTGCCCAACCATGGCTTCAAGTGG 1660
 QY 1021 GCATGAAGCTGGAGCGCTGACCTGATGGAGCGCCCGGCTCATCTGTGTGCCACCGTGA 1080
 Db 1661 GCATGAAGCTGGAGCGCTGACCTGATGGAGCGCCCGGCTCATCTGTGTGCCACCGTGA 1720
 QY 1081 AACAGTGTGTCATCGGCTCTCTAGCATTCACATTTGACGGCTGGAGCAGCAGTACGACC 1140
 Db 1721 AGCGGGTGTGTCATCGGCTCTCTAGCATTCACATTTGACGGCTGGAGCAGCAGTACGATC 1780
 QY 1141 AGTGGTGGACTGGAGTCCCGACACATCTACCCGCTCGGCTGGTGTGAGCTCACCGGCT 1200
 Db 1781 AGTGGTGGACTGGAGTCCCGACATCTACCCGCTCGGCTGGTGTGAGCTCACCGGCT 1840
 QY 1201 ACCAGTCCAGCTTCCTGTGTGCCCGCAGAACCGGCCACACCGCTGAAGGCCAAGAGGCCA 1260
 Db 1841 ACCAGTCCAGCTTCCTGTGTGCCCGCAGAACCGGCCACACCGCTGAAGGCCAAGAGGCCA 1900
 QY 1261 CAAGAAGAAAGAAACAGTTTGGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1320
 Db 1901 CAAGAAGAAAGAAACAGTTTGGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1373
 QY 1321 GACCCCTCAGACAGGGGTCCAAGAGCCCTGTGTGGAGGACGACCTTCAGGGT 1373
 Db 1961 GGCCCTCAGACAGGGGTCCAAGAGCCCTGTGTGGAGGACGACCTTCAGGGT 2013

RESULT 13
 AY237002 3725 bp mRNA linear ROD 15-APR-2003
 LOCUS Mus musculus M4MBT variant C (M4mbt) mRNA, complete cds;
 DEFINITION alternatively spliced.
 ACCESSION AY237002
 VERSION AY237002.1 GI:29569827
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 3725)
 Markus, J., Feikova, S., Sramko, M., Wolff, L. and Bies, J.
 Cloning and molecular characterization of the novel murine gene
 M4mbt encoding a nuclear zinc finger protein with four mbt domains
 Unpublished
 2 (bases 1 to 3725)
 Markus, J., Feikova, S., Sramko, M., Wolff, L. and Bies, J.
 Direct Submission
 Submitted (14-FEB-2003) Department of Molecular Virology, Cancer
 Research Institute of Slovak Academy of Sciences, Vlárska 7,
 Bratislava, Slovak Republic
 Location/Qualifiers
 1. 3725
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="Swiss Webster"
 /db_xref="taxon:10090"
 1. 3725
 /gene="M4mbt"
 46. 1227
 /note="alternatively spliced"
 /codon_start=1
 /product="M4MBT variant C"
 /protein_id="AA084918.1"
 /db_xref="GI:29569828"
 /translation="MEKPRGTETAPSPSEPEEEDLDLFGGYDSFRSYNSAGSES"

SSYLEESEAENEDREAGELPTSLHLFFSSANNRSLDGSSEPAVCMGVGTREAF
 FSKTRFCVSCSRYSNSKASITLALQKPTPKAKVLHKAASAKIGAFTHAQ
 TQGLADGPTGQDALVLFQWKGKLSYKAAVSCFKHVLVDQWEDVNMGMKEV
 LNSDAVLPSRVYVYATVIOAGYRVLLRYEGFENDASHDFWNLGTVDVHPIGWCAIN
 SKILVPRPTIHAKETDKMLKELVGSRTLPADEFHLMVESMKYPRQGMRLREVDK
 TQVSRTRMAVVDTVIGRLRLLYEDGSDDDDFWCHMMSPLIHPVGNRRVGHGKMSG
 Q"
 313. .387
 /gene="M4mbt"
 /note="Region: putative zinc finger"
 580. .894
 /gene="M4mbt"
 /note="Region: mbt repeat"
 916. .1212
 /gene="M4mbt"
 /note="Region: mbt repeat"
 BASE COUNT 879 a 991 c 1046 g 805 t 4 others
 ORIGIN

Query Match 33.9%; Score 867.4; DB 10; Length 3725;
 Best Local Similarity 71.6%; Pred. No. 3.3e-211;
 Matches 1351; Conservative 0; Mismatches 206; Indels 330; Gaps 3;

QY 1 AGCAGCTCCCACTCTATCACCAAGTGGGAGGATGTGATGAAAGGATGAAGGTGGAGTGC 60
 Db 641 AACATGTACCCCTCTATACCAAGTGGGAGAGCTCATGAAGGGATGAGTGGAGTGC 700
 QY 61 TCACAGTGTATGTGTGTCTCCCAAGCGGGTGTACTGATCGCTCTGTCTATCCAGACAG 120
 Db 701 TCACAGCGAGTGTGTGTCTCCCAAGCGGGTGTACTGATCGCTCTATCCAGGAG 760
 QY 121 CAGGGTATCGGGTGTGTGTCTCGGTATGAAGCTTTGAAATGACGCCAGCCATGACTTCT 180
 Db 761 CTGGGTACCGGTGTGTCTCGGTATGAAGCTTTGAAATGACGCCAGCTATGACTTCT 820
 QY 181 GGTCAACCTGGGAACAGTGTGATGTCCACCCCATTTGCTGTGTGCTTCAACCAAGCAAGA 240
 Db 821 GGTCAACCTGGGAACAGTGTGATGTCCACCCCATTTGCTGTGTGCTTCAACCAAGCAAGA 880
 QY 241 TCCTAGTGTCCCGCAGGACCATTCATGCCAAGTTCCAGCTGGAGGGCTACCTCATGA 300
 Db 881 TCCTGTGTACCTCCAGGACCATTCATGCCAAGTTTACTGACTGGAAGAGTACCTCATGA 940
 QY 301 AACGGCTGGTGGGTCCAGGACGCTTCCCGTGGGATTTCCACATCAAGATGTGGAGAGCA 360
 Db 941 AGCGGTGGTGGGTCCAGGACACTTCTCGAGACTTCCATATCAAGATGTGGAGAGCA 1000
 QY 361 TGAAGTACCCCTTTAGCAGGCGCATGCGGCTGGAGTGGTGGAGCAAGTCCAGGTGTAC 420
 Db 1001 TGAAGTACCCCTTTCCGACAGGCGCATGCGCTTAGAGGTTGTAGACAAGACTCAGGTGTAC 1060
 QY 421 GCACCTCGCATGGCTGTGTGGTGGACACAGTAATCGGGGGTTCGCTACGGCTCTCTACGAGG 480
 Db 1061 GTACCCGATGGCGGTGGTGGACACAGTAATCGGGGGTTCGCTACGGCTCTCTATGAGG 1120
 QY 481 ATGTTGACAGTGGACGACACTTCTGTGTGCCACATGTGGAGGCCCTGATCCACCGAGTGG 540
 Db 1121 ATGTTGACAGTGGACGACACTTCTGTGTGCCACATGTGGAGTCCCTGATCCACCGAGTGG 1180
 QY 541 GTTGGTACAGCGTGTGGGCCACCGCATCAGAGTGC----- 577
 Db 1181 GTTGGTCCCGCGTGTGGGCCACCGCATCAGAGTGC----- 577
 QY 578 ----- 577
 Db 1241 TGGGACTCCAGGTTTCAGGCTCCCGCTTAAAGATCAGAGGACCTTTTGTGTGCCCCACAGT 1300
 QY 578 ----- 577
 Db 1301 AGTGGATGAAGTGGCGTGTCTATTGCTACCTGCAAGTGAACAATGTTGAGGATCAGAGAG 1360
 QY 578 ----- 577

QY 2292 AAAAATAATTCATCCAGATTCCTTTGTAGTAAAGGTCCTAGTCTTGTAGTGGAGCTCT 2351
DB 361 AAAAATAATTCATCCAGATTCCTTTGTAGTAAAGGTCCTAGTCTTGTAGTGGAGCTCT 420
QY 2352 AGAGAGCTGGGCTGTATGCTTTTGGCCCTTTGTTC 2389
DB 421 AGAGAGCTGGGCTGTATGCTTTTGGCCCTTTGTTC 458

RESULT 15
LOCUS AX046388 3590 bp DNA linear PAT 24-NOV-2000
DEFINITION Sequence 55 from Patent WO0011168.
ACCESSION AX046388
VERSION AX046388.1 GI:11344369
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE Lemischka, I. and Moore, K.
AUTHORS Genes that regulate hematopoietic blood forming stem cells and uses
TITLE thereof
JOURNAL Patent: WO 0011168-A 55 02-MAR-2000;
Princeton University (US)
FEATURES
Location/Qualifiers
source 1..3590
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
BASE COUNT 1101 a 702 c 797 g 990 t
ORIGIN

Query Match 16.2%; Score 413.8; DB 6; Length 3590;
Best Local Similarity 60.2%; Pred. No. 8.5e-95;
Matches 735; Conservative 0; Mismatches 462; Indels 24; Gaps 2;
QY 1 AGCAGTCCCACTCTATGACAGTGGGAGATGTGTAAGGGATCAAGTGGAGTGC 60
DB 750 AGCATGACCTATGGGACCTGTGGGGTGATCTCGGAAATGTAAGGATAGAAGTTC 809
QY 61 TCAACAGTGTATGCTGTCCCAAGCGGGTGTACTGGATCGCTCTGTCAATCAGACAG 120
DB 810 CCAATACAGACTGAGCTACCTACCAAGTCTCTCGATTGCTGGAATTAATAATTAG 869
QY 121 CAGGTATCGGGTCTGCTCGGTATGAAGCTTTGAAATGAGCCAGCCATGACTTCT 180
DB 870 CAGGTATTAATGCGCTTTTGAGATATGAAGATTTGAAATGATTTCTCTCGACTTCT 929
QY 181 GGTCAACCTGGGAACAGTGTATCCACCCATTGGCTGTGTGCCATCAACAGCAAGA 240
DB 930 GTGCAATATATGTTGGTCTGTATTTATCCAGTGTGTGTGTGAGTAGTGGAAAC 989
QY 241 TCCTAGTGCCTCCAGGACCATCCATGCCAAGTTCACGATCGAAGGGTACTCTATGA 300
DB 990 CTCCTGTCTCTCTAGAACTGTTCAACATAAATATACAACTGGAAGCTTTCTAGTAA 1049
QY 301 AACGGTGTGGGCTCCAGGAGCTTCCCGTGGATTTCCACATCAGATGGTGGAGCA 360
DB 1050 AAAGACTTACTGTGGCCAAACACTTCTCTGATTTTCTCAGAAAGTTTCTGAGAGTA 1109
QY 361 TGAAGTACCCCTTTAGGCAGGGCATGGGCTGGAGTGGTGGACAGTCCAGGTGTCTAC 420
DB 1110 TGCAATATCTTTCAAACTTGTGATGAGTAGAAGTAGTTGCAAGAGGCATTATGTC 1169
QY 421 GCATCGCATGCTGTGTGTGACACAGTATCCGGGGTCCCTACGGCTCTCTACGAGG 480
DB 1170 GAAACAAGTAGTACGTGGTGAAGAGTGTAAATTGAGGACCACTACGGCTGGTATGAAG 1229
QY 481 ATGGTGCACAGTGT---AGACGACTTCTGGTGCCACATGTGGAGCCCTCTGATCCACCCAG 537
DB 1230 AGAGTGAAGATGGACAGACGACTTCTGGTGCCACATGACAGCCCTTAATCCACCATTA 1289

Search completed: February 4, 2004, 16:20:37
Job time : 9222 secs

QY 538 TGGGTTGGTTCAGCGCTGTGGGCCACCGCATCAAGATGTCTAGAGAGCGGAAGTGCATGG 597
DB 1290 TTGGATGTCTAAGAAGCATAGCCCATCGATTCAAGAGATCAGATATTACGAAGAAACAGG 1349
QY 598 CCATCACCACCCACCTTCCGGAAGATCTACTGTGATGCGGTTCTTACTCTTCTCAAGAGG 657
DB 1350 ACGGAC-----ATTTTCGATACACCTCCACACTTATTTGCTAAGG 1388
QY 658 TACGAGCAGTCTTACACAGAAGCGGTTGGTTTGAAGGAGGATGAAGCTGGAGGCCATTG 717
DB 1389 TAAAGAGTAGACACAGAGTGGAGATGGTTCAAGAGAGGAATGAAATTGAAGCTATAG 1448
QY 718 ACCCCCTGAATCTGGGCAACATCTCTGCTGGCAACTGTCTGTAAAGTCTCTCTGTGATGAT 777
DB 1449 ACCCATTAATCTTTCTTACAATATGTGTGCCACCATTAGAAAGGTGCTGGCTCATGGAT 1508
QY 778 ACTCATGATCTGTGTGACGCGGGGCCCTCCACAGATGGCTTGGACTGTGTTCTGCTACC 837
DB 1509 TCCTGATGTTGGGATTTGATGGCTCAGAGCAGCAGATGGATCTGACTGGTTCCTGTTATC 1568
QY 838 ATGCTCTTCCACGCCCATCTTCCCGGCCACTTCTGTCAAGAAATGACATTCAGTCTCA 897
DB 1569 ATGCAACCTCTCTCTCCATTTTCCCTGTGGGTTTCTGTGAAATTAACATGATAGAACTGA 1628
QY 898 CACGCCCAAGAGTTATGAGGCACAGACTTTCAACTGGGAGAACTACTTGGAGAGACCA 957
DB 1629 CTCACCCAGAGGTTTACAAAACTTCTCTTTTAAATGGTTGACTTACCTCAGGGAACCG 1688
QY 958 AGTCGAAAGCGCTCCATCGAGACTCTTTAAACATGGATTGCCCCAAACCATGGCTTCAAGG 1017
DB 1689 GCTCCATTGACAGCACCAGTAAACTATTTAATAAGGATGTTCCAAACCCAGGATTCGGT 1748
QY 1018 TGGGATGAAGCTGGAGGCCCTGACCTGATGGAGCCCGCTCATCTGTGTGSCCAGG 1077
DB 1749 TAGGAATGAAATTAAGAGCTGTAGATCTCATGGAGCCACGGTTAATATGTGTAGCCACAG 1808
QY 1078 TGAACGAGTGTGTCATCGGCTCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACG 1137
DB 1809 TTACTCGAATTTTACCCTCTCTTGAGGATACATTTTGATGGTTGGGAAGAAGTATG 1868
QY 1138 ACCAGTGGGTGAGTGGAGTCCCGACAGATCTACCCCGTGGGTGGTGTGAGTCAAG 1197
DB 1869 ACCAGTGGGTGAGTGTGAGTCCCGCTGACCTCTATCTGTAGGCTGTGTGAGTCACTG 1928
QY 1198 GCTACCGCTCCAGCCTCCTG 1218
DB 1929 GATATCAACTACAGCCTCCAG 1949